The *SERPINE2* Gene Is Associated with Chronic Obstructive Pulmonary Disease

Dawn L. DeMeo,[1,2,4] Thomas J. Mariani,[2,4] Christoph Lange,[5] Sorachai Srisuma,[2,4,10] Augusto A. Litonjua,[1,3,4] Juan C. Celedón,[1,3,4] Stephen L. Lake,[1,4] John J. Reilly,[2,4] Harold A. Chapman,[11] Brigham H. Mecham,[2,4] Kathleen J. Haley,[2,4] Jody S. Sylvia,[1] David Sparrow,[1,2,4,6,7] Avrum E. Spira,[7,8] Jennifer Beane,[8] Victor Pinto-Plata,[9] Frank E. Speizer,[1,4,5] Steven D. Shapiro,[2,4] Scott T. Weiss,[1,4,5] and Edwin K. Silverman[1,2,4]

[1]Channing Laboratory and [2]Division of Pulmonary and Critical Care Medicine, Department of Medicine, Brigham and Women's Hospital, [3]Division of Pulmonary and Critical Care Medicine, Department of Medicine, Beth Israel Deaconess Medical Center, [4]Harvard Medical School, [5]Harvard School of Public Health, [6]Veterans Affairs Medical Center, [7]The Pulmonary Center and [8]Bioinformatics Program, College of Engineering, Boston University Medical Center, and [9]Division of Pulmonary and Critical Care, Department of Medicine, Caritas-St. Elizabeth's Medical Center, Boston; [10]Department of Physiology, Faculty of Medicine, Siriraj Hospital, Mahidol University, Bangkok; and [11]Division of Pulmonary and Critical Care Medicine, Department of Medicine, University of California at San Francisco, San Francisco

Chronic obstructive pulmonary disease (COPD) is a complex human disease likely influenced by multiple genes, cigarette smoking, and gene-by-smoking interactions, but only severe alpha 1-antitrypsin deficiency is a proven genetic risk factor for COPD. Prior linkage analyses in the Boston Early-Onset COPD Study have demonstrated significant linkage to a key intermediate phenotype of COPD on chromosome 2q. We integrated results from murine lung development and human COPD gene–expression microarray studies with human COPD linkage results on chromosome 2q to prioritize candidate-gene selection, thus identifying *SERPINE2* as a positional candidate susceptibility gene for COPD. Immunohistochemistry demonstrated expression of serpine2 protein in mouse and human adult lung tissue. In family-based association testing of 127 severe, early-onset COPD pedigrees from the Boston Early-Onset COPD Study, we observed significant association with COPD phenotypes and 18 single-nucleotide polymorphisms (SNPs) in the *SERPINE2* gene. Association of five of these SNPs with COPD was replicated in a case-control analysis, with cases from the National Emphysema Treatment Trial and controls from the Normative Aging Study. Family-based and case-control haplotype analyses supported similar regions of association within the *SERPINE2* gene. When significantly associated SNPs in these haplotypic regions were included as covariates in linkage models, LOD score attenuation was observed most markedly in a smokers-only linkage model (LOD 4.41, attenuated to 1.74). After the integration of murine and human microarray data to inform candidate-gene selection, we observed significant family-based association and independent replication of association in a case-control study, suggesting that *SERPINE2* is a COPD-susceptibility gene and is likely influenced by gene-by-smoking interaction.

Chronic obstructive pulmonary disease (COPD [MIM 606963]) is a complex human disease likely influenced by multiple genes, environmental factors (especially cigarette smoking), and gene-by-smoking interactions. As the worldwide disease burden of COPD continues to increase, better treatments may be developed through understanding the variable susceptibility for developing this devastating lung disease. Cigarette smoking is the most important environmental risk factor for the development of COPD, but individuals vary widely in their susceptibility to the pulmonary effects of tobacco smoke (Burrows et al. 1977).

The only proven genetic risk factor for COPD is severe deficiency of alpha 1-antitrypsin (Laurell and Eriksson 1963). However, this genetic deficiency is present in only 1% of individuals with COPD, suggesting that COPD represents a complex disease with contributions from multiple genes and environmental risk factors. Genome-wide linkage analyses in the Boston Early-Onset COPD Study have helped to narrow the scope of candidate-gene investigation for COPD (Silverman et al. 2002*a*, 2002*b*; Palmer et al. 2003), but the linkage peaks defined still represent large genomic regions with many biologically plausible candidate genes. Candidate-gene association studies based on the known pathophysiology of COPD have been characterized by inconsistent replication of positive association results (Hersh et al. 2005). Significant linkage of an intermediate phenotype of COPD—the ratio of forced expiratory volume at one second to forced vital capacity (FEV_1/FVC)—with chromosome 2q has been observed in extended pedigrees ascertained through probands with severe, early-onset

Received August 25, 2005; accepted for publication November 17, 2005; electronically published December 15, 2005.

Address for correspondence and reprints: Dr. Dawn L. DeMeo, Channing Laboratory, 181 Longwood Avenue, Boston, MA 02115. E-mail: dawn.demeo@channing.harvard.edu

Am. J. Hum. Genet. 2006;78:253–264. 0002-9297/2006/7802-0008$15.00

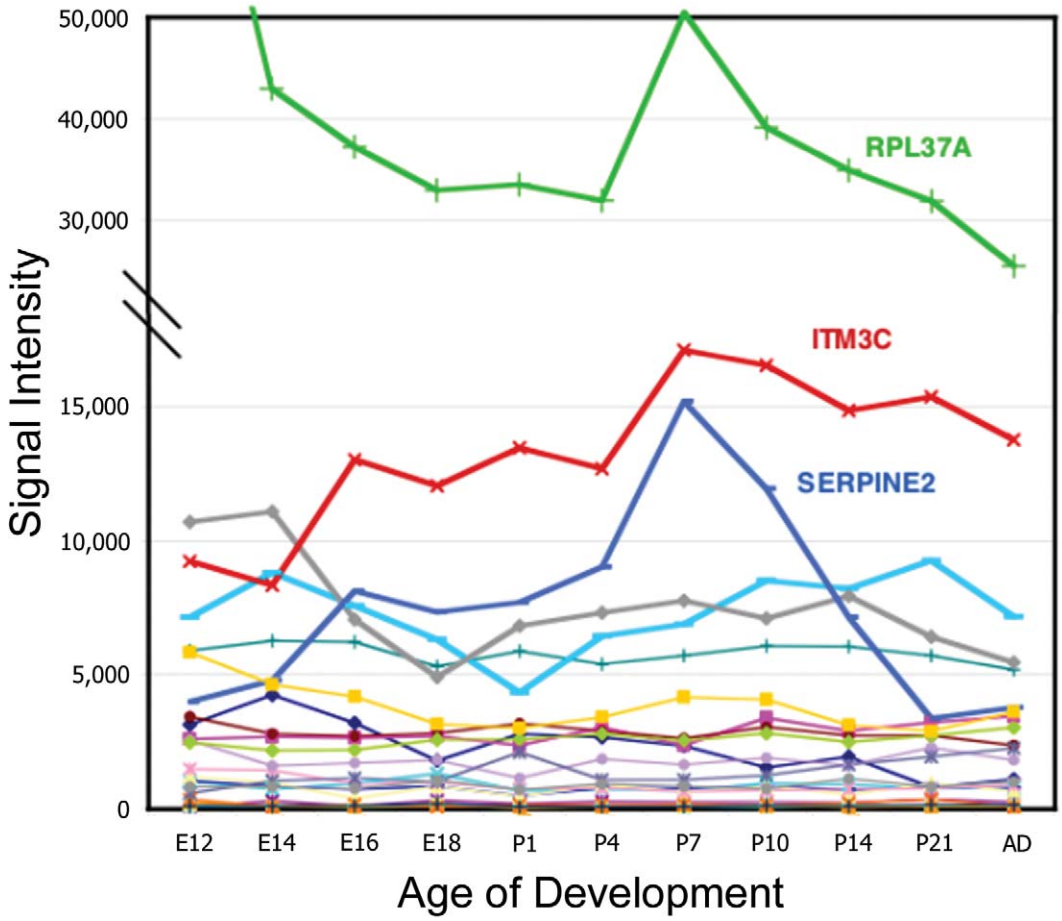

Figure 1 Gene expression during normal murine lung development. We assessed the expression of genes within the chromosome 2–linked locus, using a microarray data set of normal mouse lung development. Shown are the expression profiles for 25 probe sets, representing 23 mouse orthologues for human genes present within the locus. Signal intensities were determined using MAS 4.0. Most genes were expressed at low levels in the lung (signal intensity <4,000); seven genes were expressed at a signal intensity >4,000. Three of these genes were expressed at a signal intensity >15,000 during at least one time point, including RPL37A (a ribosomal protein), ITM2C (an integral membrane protein highly expressed in brain), and *SERPINE2*. E = embryonic day; P = postnatal day; AD = adult.

COPD in the Boston Early-Onset COPD Study (Silverman et al. 2002*b*; Palmer et al. 2003). This region of linkage likely harbors genes that contribute to COPD susceptibility through gene-by-smoking interactions (DeMeo et al. 2004). A similar region on chromosome 2q has been linked to FEV_1/FVC in the general population (Malhotra et al. 2003).

We intersected human linkage results on chromosome 2q with data from expression-array analysis of murine and human lung tissues, to prioritize the investigation of positional COPD candidate genes in this region and to identify new candidate genes for COPD. Using this methodology, we identified *SERPINE2* for further investigation as a novel candidate COPD-susceptibility gene, potentially relevant in both early-onset and later-onset COPD not related to alpha 1-antitrypsin deficiency.

Methods

Murine Expression-Array Experiments

We evaluated the expression of genes within a genomic region on chromosome 2q bracketed by positions 212,742,577 and 230,264,438 (UCSC Genome Bioinformatics); this region represents the 1.5-LOD unit of support interval for linkage on 2q (Silverman et al. 2002*b*). We evaluated the expression of genes within this genomic region in a microarray data set of normal mouse lung development, generated using the Affymetrix Mu11K platform (Mariani et al. 2002). Sequence verification of microarray probe sets was performed using the Reference Sequence database with the resources available at the Lung Transcriptome Web site, as described elsewhere (Mecham et al. 2004). Tests for significance of fold change were per-

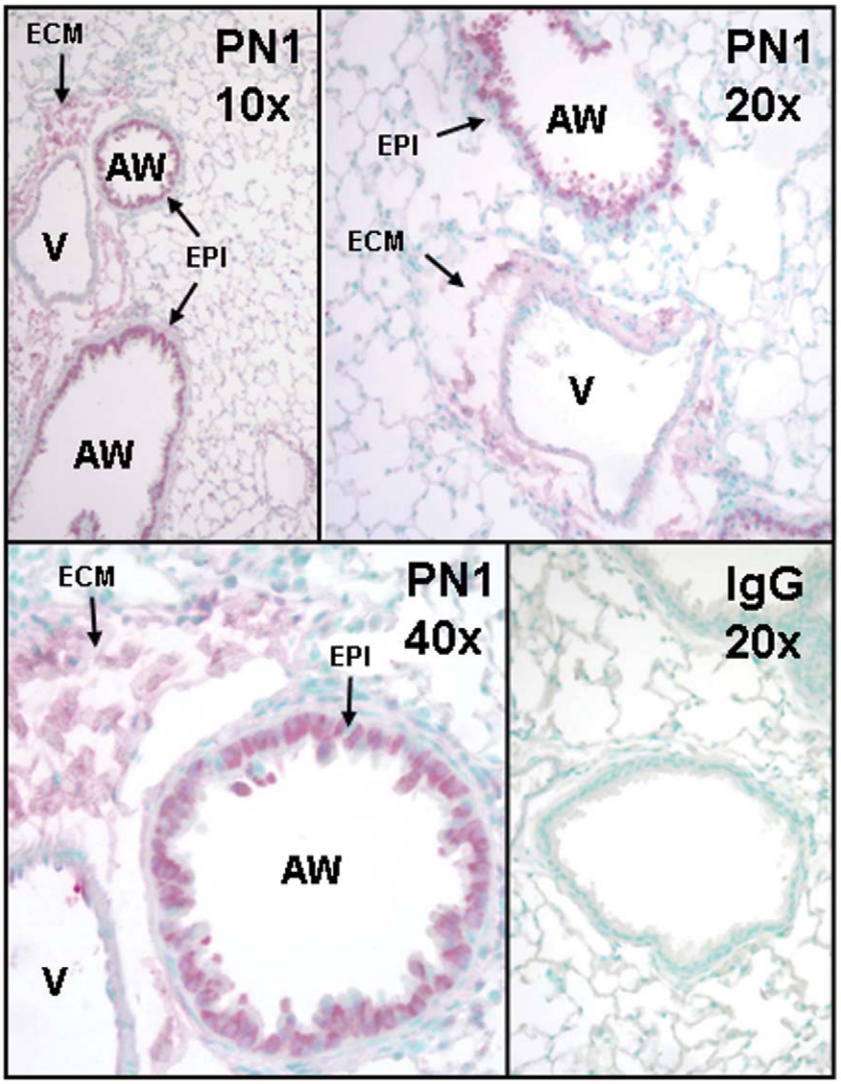

Figure 2 *SERPINE2* localization in normal adult mouse lung. Immunohistochemistry for *SERPINE2* in the normal adult mouse lung was performed using a rabbit polyclonal antibody (PN1). Staining was evident in small airway (AW)–conducting epithelium (EPI) and in the vascular (V) adventitia in an extracellular matrix-associated (ECM) pattern. No staining was observed for control IgG.

formed using the variable fold-change threshold method (Mariani et al. 2003).

Human COPD Microarray Experiments

A human microarray data set was available, generated using the Affymetrix human U133A platform; a complete description of case and control selection has been published elsewhere (Spira et al. 2004). This data set included 34 lung tissue samples; 18 were from individuals with moderate-to-severe emphysema who underwent lung volume–reduction surgery (LVRS) for COPD, and 16 were non-LVRS normal-to-minimally emphysematous control human lung samples from individuals undergoing pulmonary nodule resection. Expression values were generated from the raw image files by use of the robust microarray averaging method (Irizarry et al. 2003), including the quantile normalization procedure, as implemented in Bioconductor. Pulmonary function test information was available for these patients, including pre- and postbronchodilator FEV_1, FEV_1/FVC, total lung capacity (TLC), diffusing capacity for carbon monoxide (DLCO), and intensity of smoking (pack-years [number of packs smoked per day multiplied by the number of smoking years] of cigarettes smoked). The significance analysis of microarray (SAM) method (Tusher et al. 2001) was implemented using the publicly available TIGR Multi-Array Viewer (MEV) software (TM_4). Additionally, the Pearson and Spearman rank correlations were calculated for the association of *SERPINE2* expression with each of the lung function parameters. The significance of each correlation was determined using the Student's *t* test. Raw data from the human microarray experiments have been deposited in the Na-

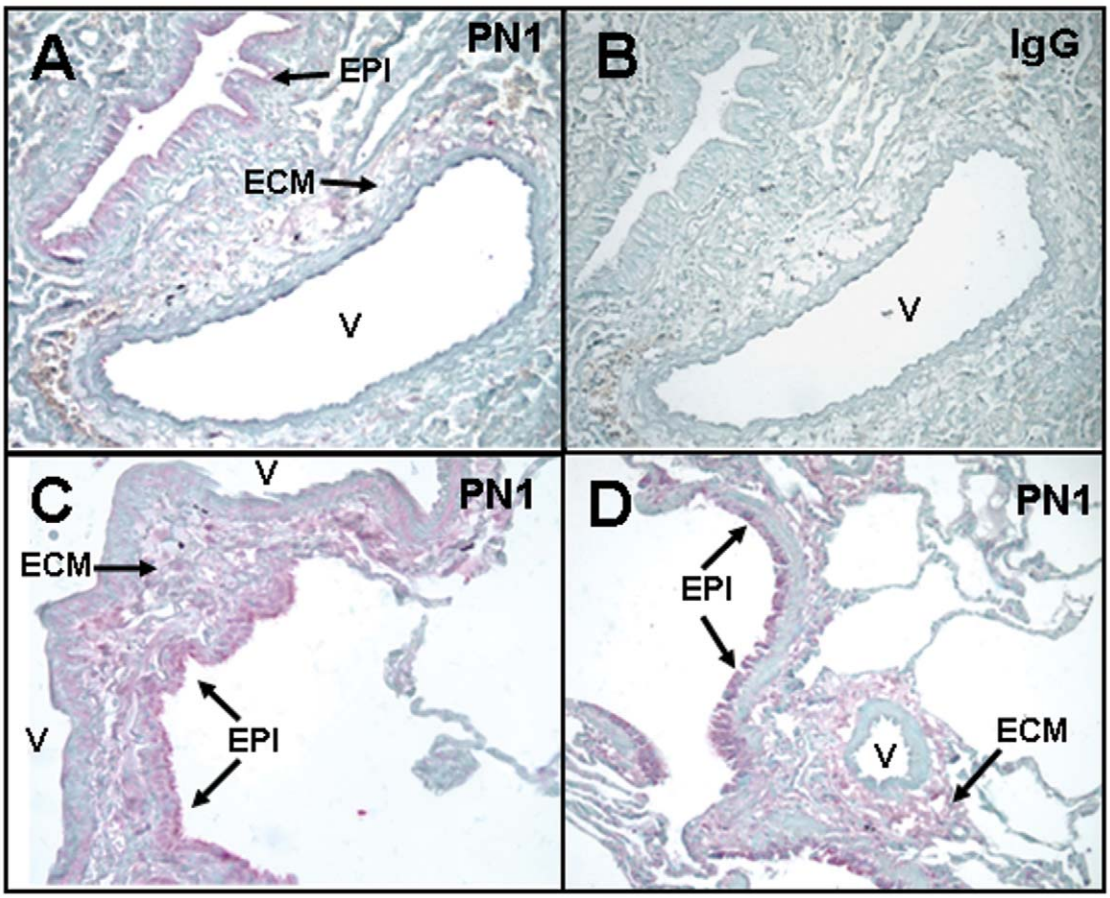

Figure 3 *SERPINE2* localization in human lung. *SERPINE2* immunostaining was observed in normal (*A*), emphysematous (*C*), and asthmatic (*D*) human lung. Staining was evident in small airway epithelial cells (EPI) and in the vascular adventitia extracellular matrix (ECM). No staining was observed for control IgG (*B*).

tional Center for Biotechnology Information (NCBI) Gene Expression Omnibus (series reference number GSE1650).

Immunohistochemistry

Immunostaining in normal adult mouse and human lung tissue was performed using a rabbit polyclonal antibody raised against recombinant bovine serpine2. The polyclonal antibodies were produced in rabbits against bovine serpine2, by use of recombinant, purified serpine2 and bovine follicular fluid. The antibodies were validated by immunoblotting. The antibodies used in our mouse and human immunohistochemistry experiments were a generous gift from Dr. J. G. Lussier. The complete method of the serpine2 antibody development has been described elsewhere (Bedard et al. 2003).

Human Populations for Genetic Association Studies

Three separate cohorts were analyzed, including participants in the Boston Early-Onset COPD Study, the National Emphysema Treatment Trial (NETT), and the Normative Aging Study (NAS). The Partners Institutional Review Board (IRB) for human studies approved the study protocols. Anonymized data sets were used for the NAS participants, as approved by the IRBs of both Partners and the Veterans Administration Hospitals.

Boston Early-Onset COPD Study

The recruitment and characteristics of probands and family members enrolled in the Boston Early-Onset COPD Study have been reported elsewhere (Silverman et al. 1998). In brief, ascertainment criteria for probands in the Boston Early-Onset COPD Study included an FEV_1 <40% of predicted, age <53 years, and no evidence of severe alpha 1-antitrypsin deficiency. All available first-degree and older second-degree relatives were invited to participate in the study; there were 127 probands and a total of 949 individuals in the family-based association analysis for *SERPINE2*.

Each study participant completed a modified version of the 1978 American Thoracic Society–Division of Lung Diseases respiratory questionnaire (Ferris 1978). Spirometry was performed as described elsewhere (Silverman et al. 1998). Most individuals completed pre- and postbronchodilator spirometry

Table 1

***SERPINE2* SNPs Genotyped and Analyzed in the Family-Based and Case-Control Association Analyses**

SNP ID	SNP	Alleles	Position	Location from ATG Start Site
1	*rs282253*	C/G	5′ Genomic/promoter	g.−39863
2	*rs1438831*	C/T	5′ Genomic/promoter	g.−39698
3	*rs7579646*	A/G	Intron 1	g.−34079
4	*rs840088*	C/T	Intron 1	g.−33227
5	*rs7562213*	A/G	Intron 1	g.−31810
6	*rs920251*	A/G	Intron 1	g.−26328
7	*rs1371029*	A/G	Intron 1	g.−21856
8	*rs2099601*	A/C	Intron 1	g.−18689
9	*rs2083120*	C/T	Intron 1	g.−18591
10	*rs2083121*	C/T	Intron 1	g.−18385
11	*rs2099602*	C/T	Intron 1	g.−18298
12	*rs4674849*	C/T	Intron 1	g.−18228
13	*rs4574111*	A/G	Intron 1	g.−17755
14	*rs1438829*	C/T	Intron 1	g.−17086
15	*rs6436459*	C/T	Intron 1	g.−14397
16	*rs1371028*	C/T	Intron 1	g.−12833
17	*rs1438828*	A/G	Intron 1	g.−11684
18	*rs1866152*	C/T	Intron 1	g.−11462
19	*rs7588220*	A/G	Intron 1	g.−6987
20	*rs1530020*	G/T	Intron 1	g.−4753
21	*rs1530021*	C/T	Intron 1	g.−4619
22	*rs6436454*	C/T	Intron 1	g.−4104
23	*rs3948261*	A/G	Intron 1	g.−3275
24	*rs2118409*	C/G	Intron 1	g.−3026
25	*rs6742903*	C/T	Intron 1	g.−2176
26	*rs12436*	A/C	Exon 2	g.41
27	*ss49785623*	A/G	Exon 2	g.190
28	*rs7581619*	A/G	Exon 2	g.230
29	*rs3795877*	C/T	Intron 2	g.440
30	*rs1866153*	A/G	Intron 2	g.2065
31	*rs6747096*	A/G	Exon 3	g.3775
32	*rs3795879*	C/T	Intron 3 exon/intron boundary	g.3796
33	*ss49785624*	A/C	Intron 3	g.3917
34	*rs6715768*	A/G	Intron 3	g.4126
35	*rs6738983*	C/T	Intron 3	g.7867
36	*rs6721140*	A/G	Intron 3	g.7974
37	*rs2076924*	C/T	Intron 3	g.9573
38	*rs6712954*	A/G	Exon 4	g.9967
39	*rs2099603*	A/G	Intron 4	g.13936
40	*rs7605945*	C/T	Intron 5	g.18811
41	*rs975278*	C/T	Intron 5	g.18910
42	*rs10164837*	C/T	Intron 6	g.19349
43	*rs729631*	C/G	Intron 7	g.21698
44	*ss49785625*	A/G	Intron 7	g.21753
45	*rs7597833*	C/T	Intron 8	g.24467
46	*ss49785626*	C/T	Intron 8	g.24583
47	*rs6734100*	C/G	Intron 8	g.24622
48	*rs1025734*	C/T	3′ Genomic	g.31248

NOTE.—Position in Human Genome Working Draft (Goldenpath hg17, dbSNP123) from UCSC Genome Bioinformatics.

(performed ~15 minutes after albuterol administration); we focused on the results from postbronchodilator trials (Palmer et al. 2003). For the analysis of quantitative phenotypes, absolute volume spirometric results with covariate adjustment, as detailed below, were used in all analyses. For the analysis of qualitative phenotypes, percent predicted values for FEV_1 and FEV_1/FVC were calculated for white adult participants by use of equations formulated by Crapo et al. (1981). For white individuals aged <18 years, predicted values were determined from equations by Hsu et al. (1979), with predicted values for FEV_1/FVC calculated from equations by Knudson et al. (1983). For African American participants, predicted values were de-

Table 2

PBAT of Quantitative Spirometric Phenotypes with *SERPINE2* SNPs, Including a SNP-by-Smoking Interaction Term

SNP	Allele	*P* by Spirometric test	
		FEV_1 Postbronchodilator	FEV_1/FVC Postbronchodilator
rs7579646	G	.0009	.0008
rs6436459	C	.009	.0005
rs1530020	A	.02	.05
rs6436454	T	.01	.02
rs3948261	T	.02	.05
rs2118409	C	.04	NS
rs3795877	A	.009	.002
rs6747096	A	.02	.001
rs3795879	C	NS	.007
rs6715768	G	.0004	.0007
rs7605945	C	.004	.0004
rs975278	T	NS	.00006
rs729631	C	NS	.001
ss49785625	A	.004	NS
rs7597833	T	.007	NS
rs6734100	G	.01	.00004

Note.—Exact *P* values are presented unless $P > .05$; each model was analyzed under the assumption of an additive mode of inheritance and adjusted for age, age^2, height, height^2, pack-years, pack-years^2, sex, and a SNP–by–pack-years interaction term. NS = not significant.

termined by equations developed by Hankinson et al. (1999). For the qualitative analysis, an FEV_1 <60% predicted in the presence of FEV_1/FVC <90% predicted defined moderate-to-severe COPD; FEV_1 <80% predicted in the presence of FEV_1/FVC <90% predicted defined mild-to-severe COPD.

NETT

The NETT is a randomized multicenter treatment trial for the investigation of outcomes in a group of individuals with COPD randomized to conventional medical therapy versus LVRS for the treatment of severe emphysema (NETT Research Group 1999). All participants had an FEV_1 of ⩽45% predicted, evidence of hyperinflation on pulmonary-function tests, and emphysema on high-resolution CT scanning of the thorax at study enrollment (NETT Research Group 1999). Of those who were enrolled in the NETT Genetics Ancillary Study and had blood samples available for genotyping, 304 white individuals were included as COPD cases in the present analysis.

NAS

The NAS is a longitudinal study of aging performed by the Veterans Administration (Bell et al. 1966). At the time of study entry, the cohort consisted of 2,280 healthy community-dwelling men from the Boston area who were aged 21–80 years old between 1961 and 1969. Since 1984, participants have completed questionnaires regarding respiratory symptoms and smoking history, and spirometric evaluation has been performed. We selected 441 white control individuals with blood samples available for genotyping; these individuals had no evidence of airflow obstruction ($FEV_1 > 80\%$ predicted and FEV_1/FVC > 90% predicted at last visit) and at least a 10–pack-year smoking history (Celedón et al. 2004).

Genotyping

Initially, seven *SERPINE2* SNPs were genotyped in the Boston Early-Onset COPD Study pedigrees, NETT cases, and NAS controls, with use of Taqman 5′ exonuclease assays, with primers obtained from Applied Biosystems (ABI). Major- and minor-allele probes were labeled with different colored fluorophores, and probe fluorescence signal detection was performed using the ABI Prism 7900 Sequence Detector System.

Subsequently, DNA sequencing was performed on all *SER-*

Table 3

PBAT of Qualitative Spirometric Phenotypes with *SERPINE2* SNPs, Including a SNP-by-Smoking Interaction Term

SNP	Allele	*P* by COPD Severity	
		Moderate to Severe	Mild to Severe
rs840088	C	.0004	NS
rs6436459	C	.0003	.002
rs1530020	A	.02	.01
rs975278	T	.03	NS
ss49785625	A	.02	.01
rs7597833	T	.03	.007
rs6734100	G	.003	NS
rs1025734	T	NS	.03

Note.—Exact *P* values are presented unless $P > .05$; each model was analyzed under the assumption of an additive mode of inheritance and adjusted for age, age^2, height, height^2, pack-years, pack-years^2, sex, and a SNP–by–pack-years interaction term.

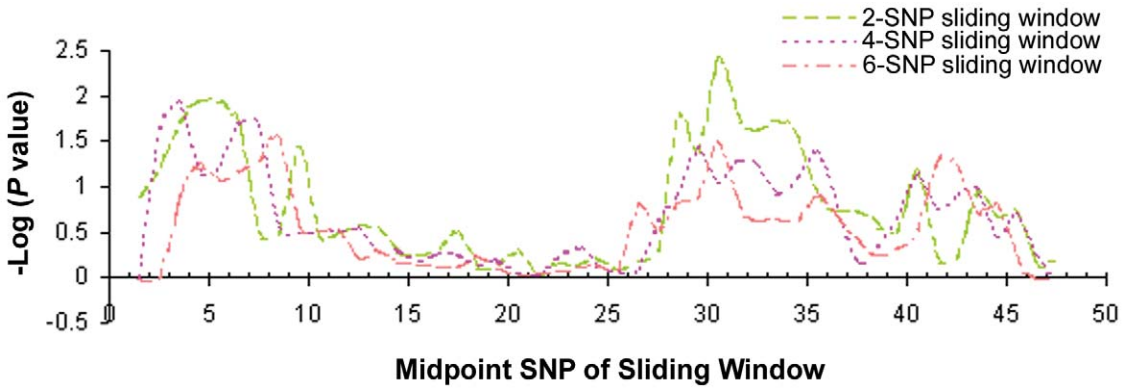

Figure 4 Simulated *P* values for sliding-window haplotype analysis of SNPs in the NETT cases and NAS controls. Graphs of the 6-, 4-, and 2-SNP sliding window–simulated *P* values (*Y*-axis) against the midpoint SNP of each sliding window (*X*-axis), demonstrate the most significant global simulated *P* value for 2-SNP sliding-window haplotypes in the regions including SNPs *rs7579646-rs840088* ($P = .02$), *rs840088-rs7562213* ($P = .01$), *rs7562213-rs920251* ($P = .01$), *rs920251-rs1371029* ($P = .02$), *rs7581619-rs3795877* ($P = .02$), *rs3795877-rs1866153* ($P = .04$), *rs1866153-rs6747096* ($P = .004$), *rs6747096-rs3795879* ($P = .02$), *rs3795879-ss49785624* ($P = .02$), *ss49785624-rs6715768* ($P = .02$), and *rs6715768-rs6738983* ($P = .02$).

PINE2 exons and exon-intron boundaries. Sequencing efforts in 34 COPD-affected probands from the Boston Early-Onset COPD Study and 12 control individuals identified 17 polymorphisms; 4 of these polymorphisms (NCBI dbSNP Build 126 ss numbers 49785623, 49785624, 49785625, and 49785626) had not been previously reported in public databases. For the second round of genotyping, 41 additional SNPs (17 identified during sequencing and 24 more from public databases) were genotyped in both the Boston Early-Onset COPD Study families and NETT-NAS case-control populations, with use of the SEQUENOM platform, for a total of 48 SNPs in both cohorts (with 7 from the first round and 41 from the second round of genotyping).

Statistical Methods for Genetic Association Analysis

Pedigree-Based Association Test (PBAT).—For the family-based single-SNP association analyses and the family-based haplotype analyses of *SERPINE2* in the Boston Early-Onset COPD Study, we used PBAT (Lange et al. 2004) (C.L. Web site). We had no biological data supporting a dominant or recessive model, so we used the additive model to minimize multiple statistical testing. The association analysis was adjusted for age, sex, height, smoking history (yes/no), and pack-years of cigarette smoking and included a SNP–by–pack-years (gene-by-environment) interaction term. Because *SERPINE2* is located in a genomic region linked to COPD-related phenotypes, the null hypothesis of our PBAT analyses was no association in the presence of linkage. In this setting, PBAT adjusts for linkage by using the empirical variance in the family-based association test (FBAT) statistic. *P* values $\leq .05$ are reported. For the models including the interaction term, a composite *P* value that includes the association for the main SNP effect and the gene-by-smoking interaction is reported. In general, optimal model fitting is possible in FBAT, by use of the conditional-power approach as described by Lange and Laird (2002). However, model selection that is based on conditional power calculation is computationally feasible only in the setting of small pedigrees, and this approach was not feasible for the extended pedigrees of the Boston Early-Onset COPD Study.

Conditional linkage models were analyzed, including the SNP data and a SNP-by-smoking interaction term by use of a variance component approach, as implemented in the Sequential Oligogenic Linkage Analysis Routines (SOLAR), version 1.7.4 (Almasy and Blangero 1998). Overall and smokers-only linkage models were generated using a panel of 377 STR markers with relevant covariates (including age, sex, height, pack-years of smoking, and *SERPINE2* SNP of interest), as described elsewhere (DeMeo et al. 2004).

Case-Control Replication Analysis.—Association between *SERPINE2* SNPs and severe COPD among NETT cases and NAS controls was investigated with the SAS Genetics program (SAS Institute). Hardy-Weinberg equilibrium was evaluated among the controls for each SNP by use of the goodness-of-fit test, as implemented through SAS Genetics. Association between each of the individual SNPs and COPD was investigated using 2×2 contingency tables and by a trend test to investigate allelic additive effects. To investigate allelic additive effects with the Cochran-Armitage trend test and to investigate dominance effects with the genotype case-control test, 2×3 contingency tables were used. Since the control population was all male, a males-only analysis was performed in addition to an analysis that included all of the NETT case participants.

Haplotype frequencies were estimated, and the association between haplotypes with a frequency >5% and COPD was tested using global and haplotype-specific statistics with the haplo.stats program (Schaid et al. 2002). Sliding windows of 6, 4, and 2 adjacent SNPs were used to test for association across the *SERPINE2* gene and to localize the region of most-significant association (Clayton and Jones 1999; Schaid et al. 2002). Simulated *P* values were also generated for 6-, 4-, and 2-SNP sliding windows by use of 1,000 replicates; the negative

log of the *P* value was graphed against the midpoint SNP of the sliding window to demonstrate the region of most-significant association.

Results

Mouse Expression-Array Experiments

In a microarray data set describing normal mouse lung development generated using the Affymetrix Mu11K platform (Mariani et al. 2002), 34 probe sets in the murine data set that matched 24 independent mouse orthologues for human genes were present within the region of human chromosome 2q linkage. Expression data from the 25 Reference Sequence–verified probe sets, representing 23 genes, are presented in figure 1. The majority of the genes were expressed in the lung at a signal intensity <4,000. Three genes were expressed at a signal intensity >15,000 during at least one time point, including RPL37A (a ribosomal protein), ITM2C (an integral membrane protein highly expressed in the brain), and *SERPINE2*. *SERPINE2* was most highly expressed during murine alveogenesis (postnatal days 4–10) and had the greatest expression change (4.5-fold) across the developmental time series. This fold change was significant with use of the variable fold-change threshold method, suggesting that this gene is regulated during normal lung development.

Human Microarray Results

A human lung microarray data set (Spira et al. 2004) from individuals with severe COPD and control subjects was evaluated for disease-associated changes in SERPINE2, ITM3C, and RPL37A expression (in this region, the three most highly expressed genes during murine lung development), by use of pulmonary-function measurements as continuous variables. *SERPINE2* was positively correlated with higher TLC (Spearman correlation = 0.46; $P = .008$, Pearson correlation = 0.42, $P = .020$); this suggested a significant correlation with lung hyperinflation. Additionally, *SERPINE2* expression was inversely correlated with postbronchodilator FEV_1 (Spearman correlation = −0.43; $P = .020$; Pearson correlation = −0.38; $P = .043$) and was inversely correlated with DLCO (Spearman correlation = −0.41; $P = .025$; Pearson correlation = −0.41; $P = .023$). Higher TLC, reduced postbronchodilator FEV_1, and reduced DLCO are all clinical features observed in patients with COPD. ITM2C and RPL37A each had a single Reference Sequence–verified probe set; neither showed significant correlation to any of the pulmonary-function measurements. Examination of these same data with use of SAM further suggested that *SERPINE2* expression was higher in individuals with severe emphysema than in controls (fold change of 1.25, with a false-discovery rate of 6.2%). This modest fold change was verified using quantitative RT-PCR. Because of the expression pattern of *SERPINE2* across normal lung development and the consistent correlations with COPD-related phenotypes, we selected *SERPINE2* for further investigation.

Immunohistochemistry

We performed immunohistochemistry for *SERPINE2* in postnatal mouse lungs and demonstrated prominent immunolocalization of *SERPINE2* in a cell-associated pattern within conducting airway epithelial cells and in an extracellular matrix-associated pattern in the vascular adventitia (fig. 2). Immunohistochemistry of adult human lungs demonstrated an analogous staining pattern (fig. 3).

PBAT

We next investigated the genetic association of a panel of 48 *SERPINE2* SNPs with COPD phenotypes (table 1). Multivariate models were evaluated in 127 pedigrees (949 individuals) in the Boston Early-Onset COPD Study. In additive models that included an interaction term to capture SNP-by-smoking (gene-by-environment) effects, 18 SNPs in *SERPINE2* demonstrated significant association with quantitative and/or qualitative spirometric phenotypes (tables 2 and 3).

Case-Control Association Testing

We attempted to replicate the family-based associations for *SERPINE2* in a case-control analysis using 304 NETT individuals with severe smoking-related COPD and 441 NAS controls, all smokers, with normal spirometry (Celedón et al. 2004). As reported elsewhere, there was no significant evidence of population stratification in these cases and controls (Celedón et al. 2004). All *SERPINE2* SNPs were in Hardy-Weinberg equilibrium in the controls. In the case-control analysis (table 4), eight SNPs (*rs1438831, rs7579646, rs840088, rs7562213, rs920251, rs3795877, rs6747096,* and *rs3795879*) demonstrated significant association ($P \leq .05$), including five SNPs that were significant in the family-based analysis. In a males-only analysis, the association remained robust.

Haplotype Analysis

Using a sliding-window approach, we analyzed adjacent 6-, 4-, and 2-SNP haplotypes and observed global significance for the 6-, 4-, and 2-SNP haplotypes (fig. 4); this significance was confirmed with empirical *P* values determined through 1,000 simulations. We used these results to narrow our individual haplotype analyses to these regions. With focus on the most-significant 2-

Table 4

Replication of Significant Associations of *SERPINE2* with COPD in the NETT Cases and NAS Controls

SNP and Sample	Allele and Frequency		Genotype and Frequency	*P* for		
				Allele Test	Genotype Test	Males-Only Genotype Test
rs1438831 (5′ genomic/promoter):	C	T	CC/CT/TT	.098	.089	***.044***
Cases	.70	.30	.48/.41/.09			
Controls	.65	.35	.41/.49/.10			
rs7579646 (intron 1):	A	G	AA/AG/GG	***.043***	.088	***.016***
Cases	.17	.83	.03/.24/.73			
Controls	.21	.79	.04/.35/.61			
rs840088 (intron 1):	C	T	CC/CT/TT	***.007***	***.009***	***.008***
Cases	.70	.30	.51/.37/.11			
Controls	.63	.37	.38/.50/.13			
rs7562213 (intron 1):	A	G	AA/AG/GG	***.020***	***.018***	***.008***
Cases	.30	.70	.12/.36/.52			
Controls	.36	.64	.11/.50/.39			
rs920251 (intron 1):	A	G	AA/AG/GG	***.015***	***.011***	***.007***
Cases	.30	.69	.11/.39/.50			
Controls	.36	.64	.12/.49/.39			
rs3795877 (intron 2):	A	G	AA/AG/GG	***.019***	***.053***	***.005***
Cases	.82	.18	.72/.25/.04			
Controls	.77	.23	.58/.37/.05			
rs6747096 (exon 3):	A	G	AA/AG/GG	***.002***	***.005***	***.0006***
Cases	.83	.17	.73/.24/.04			
Controls	.76	.24	.57/.38/.05			
rs3795879 (exon 3/intron 3 boundary):	C	T	CC/CT/TT	***.009***	***.014***	***.0007***
Cases	.81	.19	.71/.24/.04			
Controls	.75	.25	.55/.40/.05			

Note.—Only SNPs with significant results ($P \leq .05$) in any test are shown. Significant results are shown in bold italics.

SNP haplotype-specific associations, significance levels for haplotype-specific *P* values for FEV_1/FVC in the Boston Early-Onset COPD pedigrees were similar to associated haplotypes in the NETT-NAS case-control analysis (for 2-SNP haplotype *rs1438831-rs7579646*, alleles TA, family $P = .003$, case-control $P = .02$; for 2-SNP haplotype *rs1866153-rs6747096,* alleles CG, family $P = .01$, case-control $P = .001$; for SNP haplotype *rs6747096-rs3795879*, alleles GT, family $P = .047$, case-control $P = .006$).

Conditional Linkage Models

Linkage models analyzed in SOLAR, including as covariates those SNPs significant in the single-SNP and haplotype analyses, revealed attenuation of LOD scores for linkage to postbronchodilator FEV_1/FVC on chromosome 2q. Specifically, inclusion of SNP *rs7579646* in the overall linkage model resulted in attenuation of the LOD score from 4.41 (222 cM) to 2.16 (220 cM); linkage analysis of the smokers-only model resulted in further attenuation of the LOD score to 1.70. In this region of the gene, attenuation of the LOD score was also observed for SNPs *rs840088* (LOD attenuation to 2.25 in smokers-only model) and *rs7562213* (LOD attenuation to 2.07 smokers-only model). Farther downstream, LOD attenuation was observed with the inclusion of *rs3795877* (LOD attenuation to 2.22 in smokers-only model), *rs6747096* (LOD attenuation to 3.82 in smokers-only model), and *rs3795879* (LOD attenuation to 2.75 in smokers-only model).

Discussion

Collaborative approaches are crucial for identifying candidate genes and uncovering potential new pathways of complex-disease pathogenesis. The phenotypic expression of COPD is likely under the influence of multiple genes, cigarette smoking (the major environmental risk factor for COPD), and gene-by-smoking interactions. In our study, expression-array analyses were utilized for can-

didate-gene prioritization in a broad linkage region on chromosome 2q; the intersection of genomic expression-array and linkage analyses led to the identification of *SERPINE2* as a potential COPD susceptibility gene. Given our current state of understanding of COPD pathophysiology, *SERPINE2* is not an obvious COPD candidate gene.

The *SERPINE2* gene (GenBank reference sequence NM_006216) consists of nine exons, encoding a 44-kDa cellular and extracellular matrix-associated serine protease inhibitor, mainly involved in coagulation and fibrinolysis (Baker et al. 1980, 1982). The major known role of *SERPINE2* is as an inhibitor of thrombin, urokinase, and plasmin (Baker et al. 1980; Scott et al. 1985). Although *SERPINE2* has been extensively investigated in the brain and has been observed to hinder neuron apoptosis and injury-mediated cell death (Houenou et al. 1995; Rossignol et al. 2004), there have been no studies investigating *SERPINE2* in the human lung. *SERPINE2* expression has been observed in the embryonic mouse lung within the conducting-airway epithelium (Mansuy et al. 1993); our observation of expression in postembryonic mouse and human lungs supported further investigation of the role of *SERPINE2* in lung disease.

The strengths of our current approach include the identification of a plausible positional candidate gene for COPD by expression analysis and immunohistochemistry and by demonstration of significant association of multiple SNPs in *SERPINE2* in extended pedigrees, replication in an independent case-control study, and significance at both the single-SNP and haplotype levels. Although multiple statistical testing may yield false-positive associations, independent replication suggests that these are likely valid associations. LOD score attenuation in overall and smoking-stratified linkage models is further support of *SERPINE2* as a COPD candidate gene.

The data from the microarray experiments guided further investigation, hypothesis generation, and candidate-gene selection. One weakness of this approach is that expression changes in the lung are subject to variations in sampling that are unavoidable in a complex and heterogeneous tissue, and gene expression changes detected in end-stage tissue may not be restricted to those involved in COPD mechanisms (and may be secondary or compensatory responses). Although the human control samples may not have had completely normal lungs, the combination of the microarray data and immunohistochemistry supported our further investigation of *SERPINE2*. Weaknesses of our association studies include the potential heterogeneity of disease phenotypes between our different human populations. However, limiting the potential heterogeneity is the fact that both the Boston Early-Onset COPD probands and the NETT cases represent individuals with very severe COPD. The replicated association of multiple SNPs in *SERPINE2* suggests that the next phase of investigation of *SERPINE2* in the lung should include association analysis of a comprehensive set of *SERPINE2* SNPs and assessment of potential functional variants, as well as the consideration of fibrinolytic pathways in the pathogenesis of COPD. We have not observed complete attenuation of our LOD scores to zero in conditional linkage models; this suggests that other variants in the *SERPINE2* gene, or more likely variants in another COPD-susceptibility gene in the region, contribute to the linkage signal in this region on chromosome 2q.

A mechanism through which *SERPINE2* may contribute to the development of COPD has yet to be identified. To date, the expression of *SERPINE2* has been most extensively studied in the brain and reproductive system; as such, it is a surprising candidate for association with COPD. However, SERPINE2 does belong to the serpin family of proteins, as does alpha 1-antitrypsin, deficiency of which is the only known genetic cause of COPD. Our results suggest that overexpression of *SERPINE2,* rather than deficiency, is associated with COPD. Nonetheless, the aggregate results of family-based and case-control single-SNP and haplotype association analyses, as well as the conditional linkage results demonstrating LOD score attenuation, suggest the strongest significance of variants in or in linkage disequilibrium with intron 1 and exon 3 of the *SERPINE2* gene.

Alterations in coagulation and fibrinolytic pathways have been associated with acute and chronic lung injury and airway hyperresponsiveness, but extensive investigation in the setting of COPD has yet to be performed (Chambers 2003; Idell 2003; Wagers et al. 2004). Although there has been occasional speculation in the past regarding the role of vascular thrombosis and hypercoaguability in the development of emphysema, this has not been widely investigated (Brantigan et al. 1966; Alessandri et al. 1994; Ashitani et al. 2002). *SERPINE2* has been demonstrated to be an inhibitor of trypsinlike serine proteases (such as thrombin, trypsin, plasmin, and urokinase); *SERPINE2* has not been demonstrated to inhibit neutrophil elastase or chymotrypsinlike proteases (Scott et al. 1985). In addition to proteolytic stress, imbalance in oxidant-antioxidant pathways are potentially important in the development of COPD. To date, we are not aware of any studies of oxidative regulation of *SERPINE2* activity.

There exists a complex mechanism of cross talk between serine proteases (and their inhibitors) and metalloproteinases (and their inhibitors), particularly within the lung and in relation to the pathogenesis of emphysema, a major component of COPD. For instance, the serine protease neutrophil elastase is capable of degrading metalloproteinase inhibitors (TIMPs), whereas the metalloproteinase macrophage elastase (MMP-12) is capable of degrading the neutrophil elastase inhibitor al-

pha 1-antitrypsin (Shapiro et al. 2003). Plasmin, one of the potential substrates for *SERPINE2*, is also capable of contributing to metalloproteinase activation. Polymorphic variants in the *SERPINE2* gene could contribute to the development of COPD through alterations in matrix metalloproteinase pathways. The physiological substrate for *SERPINE2* within the lung and the impact of alterations in expression on the complex coordination of proteolytic activities and imbalances in the exposure to cigarette smoke remain to be defined.

In summary, through the intersection of murine and human microarray analyses with linkage information on chromosome 2q, we have demonstrated replicated association of *SERPINE2* variants with COPD. *SERPINE2* may have a role in defining susceptibility phenotypes of COPD by influencing inflammatory responses to environmental exposures such as cigarette smoke through gene-by-smoking interactions. *SERPINE2* has been demonstrated to inhibit extracellular matrix destruction (Bergman et al. 1986), which suggests that polymorphisms in the *SERPINE2* gene may influence alterations in repair of smoking-induced lung damage. More research is needed to identify critical variants in the gene as well as to define a functional role for *SERPINE2* in COPD.

Acknowledgments

We thank the study participants and their families for their enthusiastic support. We also thank Dr. J. Drazen and Dr. L. Ginns for their role in helping to establish the Boston Early-Onset COPD cohort. The authors acknowledge Dr. J. Brody and Dr. B. Celli, who were instrumental in the human microarray study. We also thank Soumyaroop Bhattacharya and Jake Abernethy for assistance with the microarray data analyses. We also thank Dr. Lussier for his generous gift of the *SERPINE2* antibody. This work has been funded by National Institutes of Health (NIH) grant K08 HL072918 (to D.L.D.); an American Lung Association Research Award (to D.L.D.); NIH grants HL071885 (to T.M.), HL61575, HL71393, and HL075478 (to E.K.S.), HL54853 (to S.D.S.), and HL67204 (to H.A.C.); and an American Lung Association Career Investigator Award (to E.K.S.). The NAS is supported by the Cooperative Studies Program/Epidemiology Research and Information Center (ERIC) of the U.S. Department of Veterans Affairs and is a component of the Massachusetts Veterans ERIC. The Human Microarray work was supported by a Doris Duke Charitable Foundation Clinical Scientist Development Award (to A.S.) and NIH grant HL71771 (to J. Brody). NETT was supported by National Heart, Lung, and Blood Institute contracts N01HR76101–N01HR76116, N01HR76118, and N01HR76119; the Centers for Medicare and Medicaid Services; and the Agency for Healthcare Research and Quality.

Web Resources

The accession number and URLs for data presented herein are as follows:

Bioconductor, http://www.bioconductor.org/
C.L. Web site, http://biosun1.harvard.edu/~clange/pbat.htm (for PBAT)
GenBank, http://www.ncbi.nlm.nih.gov/Genbank/ (for *SERPINE2* [accession number NM_006216])
Lung Transcriptome, http://lungtranscriptome.bwh.harvard.edu/
Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.gov/Omim/ (for COPD)
TM_4, http://www.tigr.org/software/tm4/mev.html (for TIGR Multi-Array Viewer [MEV] software)
UCSC Genome Bioinformatics, http://genome.ucsc.edu/ (for the June 2002 freeze)

References

Alessandri C, Basili S, Violi F, Ferroni P, Gazzaniga PP, Cordova C, Chronic Obstructive Bronchitis and Haemostasis Group (1994) Hypercoagulability state in patients with chronic obstructive pulmonary disease. Thromb Haemost 72:343–346

Almasy L, Blangero J (1998) Multipoint quantitative-trait linkage analysis in general pedigrees. Am J Hum Genet 62:1198–1211

Ashitani J, Mukae H, Arimura Y, Matsukura S (2002) Elevated plasma procoagulant and fibrinolytic markers in patients with chronic obstructive pulmonary disease. Intern Med 41:181–185

Baker JB, Low DA, Eaton DL, Cunningham DD (1982) Thrombin-mediated mitogenesis: the role of secreted protease nexin. J Cell Physiol 112:291–297

Baker JB, Low DA, Simmer RL, Cunningham DD (1980) Protease-nexin: a cellular component that links thrombin and plasminogen activator and mediates their binding to cells. Cell 21:37–45

Bedard J, Brule S, Price CA, Silversides DW, Lussier JG (2003) Serine protease inhibitor-E2 (SERPINE2) is differentially expressed in granulosa cells of dominant follicle in cattle. Mol Reprod Dev 64:152–165

Bell B, Rose CL, Damon A (1966) The Veterans Administration longitudinal study of healthy aging. Gerontologist 6:179–184

Bergman BL, Scott RW, Bajpai A, Watts S, Baker JB (1986) Inhibition of tumor-cell-mediated extracellular matrix destruction by a fibroblast proteinase inhibitor, protease nexin I. Proc Natl Acad Sci USA 83:996–1000

Brantigan OC, Kress MB, Goco RB (1966) Pulmonary emboli: a factor in the etiology and pathogenesis of pulmonary emphysema. Dis Chest 49:491–501

Burrows B, Knudson RJ, Cline MG, Lebowitz MD (1977) Quantitative relationships between cigarette smoking and ventilatory function. Am Rev Respir Dis 115:195–205

Celedón JC, Lange C, Raby BA, Litonjua AA, Palmer LJ, DeMeo DL, Reilly JJ, Kwiatkowski DJ, Chapman HA, Laird N, Sylvia JS, Hernandez M, Speizer FE, Weiss ST, Silverman EK (2004) The transforming growth factor-β1 (TGFB1) gene is associated with chronic obstructive pulmonary disease (COPD). Hum Mol Genet 13:1649–1656

Chambers RC (2003) Role of coagulation cascade proteases in lung repair and fibrosis. Eur Respir J Suppl 44:33S–35S

Clayton D, Jones H (1999) Transmission/disequilibrium tests for extended marker haplotypes. Am J Hum Genet 65:1161–1169

Crapo RO, Morris AH, Gardner RM (1981) Reference spirometric values using techniques and equipment that meet ATS recommendations. Am Rev Respir Dis 123:659–664

DeMeo DL, Celedón JC, Lange C, Reilly JJ, Chapman HA, Sylvia JS, Speizer FE, Weiss ST, Silverman EK (2004) Genome-wide linkage of forced mid-expiratory flow in chronic obstructive pulmonary disease. Am J Respir Crit Care Med 170:1294–1301

Ferris BG (1978) Epidemiology Standardization Project (American Thoracic Society). Am Rev Respir Dis 118:1–120

Hankinson JL, Odencrantz JR, Fedan KB (1999) Spirometric reference

values from a sample of the general U.S. population. Am J Respir Crit Care Med 159:179–187

Hersh CP, DeMeo DL, Lange C, Litonjua AA, Reilly JJ, Kwiatkowski D, Laird N, Sylvia JS, Sparrow D, Speizer FE, Weiss ST, Silverman EK (2005) Attempted replication of reported chronic obstructive pulmonary disease candidate gene associations. Am J Respir Cell Mol Biol 33:71–78

Houenou LJ, Turner PL, Li L, Oppenheim RW, Festoff BW (1995) A serine protease inhibitor, protease nexin I, rescues motoneurons from naturally occurring and axotomy-induced cell death. Proc Natl Acad Sci USA 92:895–899

Hsu KH, Jenkins DE, Hsi BP, Bourhofer E, Thompson V, Tanakawa N, Hsieh GS (1979) Ventilatory functions of normal children and young adults—Mexican-American, white, and black. I. Spirometry. J Pediatr 95:14–23

Idell S (2003) Coagulation, fibrinolysis, and fibrin deposition in acute lung injury. Crit Care Med 31:S213–S220

Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4:249–264

Knudson RJ, Lebowitz MD, Holberg CJ, Burrows B (1983) Changes in the normal maximal expiratory flow-volume curve with growth and aging. Am Rev Respir Dis 127:725–734

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

Lange C, Laird NM (2002) Power calculations for a general class of family-based association tests: dichotomous traits. Am J Hum Genet 71:575–584

Laurell C-B, Eriksson S (1963) The electrophoretic α_1-globulin pattern of serum in α_1-antitrypsin deficiency. Scand J Clin Lab Invest 15: 132–140

Malhotra A, Peiffer AP, Ryujin DT, Elsner T, Kanner RE, Leppert MF, Hasstedt SJ (2003) Further evidence for the role of genes on chromosome 2 and chromosome 5 in the inheritance of pulmonary function. Am J Respir Crit Care Med 168:556–561

Mansuy IM, van der Putten H, Schmid P, Meins M, Botteri FM, Monard D (1993) Variable and multiple expression of protease nexin-1 during mouse organogenesis and nervous system development. Development 119:1119–1134

Mariani TJ, Budhraja V, Mecham BH, Gu CC, Watson MA, Sadovsky Y (2003) A variable fold change threshold determines significance for expression microarrays. FASEB J 17:321–323

Mariani TJ, Reed JJ, Shapiro SD (2002) Expression profiling of the developing mouse lung: insights into the establishment of the extracellular matrix. Am J Respir Cell Mol Biol 26:541–548

Mecham BH, Klus GT, Strovel J, Augustus M, Byrne D, Bozso P, Wetmore DZ, Mariani TJ, Kohane IS, Szallasi Z (2004) Sequence-matched probes produce increased cross-platform consistency and more reproducible biological results in microarray-based gene expression measurements. Nucleic Acids Res 32:e74

NETT Research Group (1999) Rationale and design of the National Emphysema Treatment Trial (NETT): a prospective randomized trial of lung volume reduction surgery. J Thorac Cardiovasc Surg 118: 518–528

Palmer LJ, Celedón JC, Chapman HA, Speizer FE, Weiss ST, Silverman EK (2003) Genome-wide linkage analysis of bronchodilator responsiveness and post-bronchodilator spirometric phenotypes in chronic obstructive pulmonary disease. Hum Mol Genet 12:1199–1210

Rossignol P, Ho-Tin-Noe B, Vranckx R, Bouton MC, Meilhac O, Lijnen HR, Guillin MC, Michel JB, Angles-Cano E (2004) Protease nexin-1 inhibits plasminogen activation-induced apoptosis of adherent cells. J Biol Chem 279:10346–10356

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

Scott RW, Bergman BL, Bajpai A, Hersh RT, Rodriguez H, Jones BN, Barreda C, Watts S, Baker JB (1985) Protease nexin: properties and a modified purification procedure. J Biol Chem 260:7029–7034

Shapiro SD, Goldstein NM, Houghton AM, Kobayashi DK, Kelley D, Belaaouaj A (2003) Neutrophil elastase contributes to cigarette smoke-induced emphysema in mice. Am J Pathol 163:2329–2335

Silverman EK, Chapman HA, Drazen JM, Weiss ST, Rosner B, Campbell EJ, O'Donnell WJ, Reilly JJ, Ginns L, Mentzer S, Wain J, Speizer FE (1998) 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 157:1770–1778

Silverman EK, Mosley JD, Palmer LJ, Barth M, Senter JM, Brown A, Drazen JM, Kwiatkowski DJ, Chapman HA, Campbell EJ, Province MA, Rao DC, Reilly JJ, Ginns LC, Speizer FE, Weiss ST (2002*a*) Genome-wide linkage analysis of severe, early-onset chronic obstructive pulmonary disease: airflow obstruction and chronic bronchitis phenotypes. Hum Mol Genet 11:623–632

Silverman EK, Palmer LJ, Mosley JD, Barth M, Senter JM, Brown A, Drazen JM, Kwiatkowski DJ, Chapman HA, Campbell EJ, Province MA, Rao DC, Reilly JJ, Ginns LC, Speizer FE, Weiss ST (2002*b*) Genomewide linkage analysis of quantitative spirometric phenotypes in severe early-onset chronic obstructive pulmonary disease. Am J Hum Genet 70:1229–1239

Spira A, Beane J, Pinto-Plata V, Kadar A, Liu G, Shah V, Celli B, Brody JS (2004) Gene expression profiling of human lung tissue from smokers with severe emphysema. Am J Respir Cell Mol Biol 31:601–610

Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci USA 98:5116–5121

Wagers SS, Norton RJ, Rinaldi LM, Bates JH, Sobel BE, Irvin CG (2004) Extravascular fibrin, plasminogen activator, plasminogen activator inhibitors, and airway hyperresponsiveness. J Clin Invest 114: 104–111