

Secreted Modular Calcium-binding Protein 2 Haplotypes Are Associated with Pulmonary Function

Jemma B. Wilk¹, Alan Herbert², Christina M. Shoemaker³, Daniel J. Gottlieb⁴, and Samer Karamohamed^{3,5}

¹Departments of Neurology and Medicine, ²Department of Genetics and Genomics, ³Framingham Heart Study Genetics Laboratory, Department of Neurology, and ⁴Pulmonary Center, Boston University School of Medicine, Boston, Massachusetts; and ⁵Department of Human Genetics, University of Chicago, Chicago, Illinois

Rationale: Previously reported linkage to FEV₁ (LOD score = 5.0) on 6q27 in the Framingham Heart Study (FHS) led us to explore a candidate gene, *SMOC2*, at 168.6 Mb.

Objectives: We tested association between *SMOC2* polymorphisms and FEV₁ and FVC in unrelated FHS participants.

Methods: Twenty single-nucleotide polymorphisms (SNPs) around *SMOC2* were genotyped in 1,734 subjects.

Measurements and Main Results: SNP data were analyzed using multiple linear regression models incorporating sex, age, body mass index, height, and smoking history as covariates, and analyses were repeated within strata of ever- and never-smokers. The minor allele of SNP rs1402 was associated with higher mean FEV₁ ($p = 0.003$) and FVC ($p = 0.02$) measures. In never-smoking subjects, association with higher measures was observed with the minor allele of rs747995 (FEV₁, $p = 0.0006$; FVC, $p = 0.0008$). These two SNPs lie in different haplotype blocks and reside in intron 4 of *SMOC2*. Haplotype analysis revealed a common G-T haplotype (rs747995–rs1402) with 77% frequency in never-smoking FHS subjects. The G-T haplotype was associated with reduction of 126 ml for FEV₁ ($p = 0.0002$) and 157 ml for FVC ($p = 0.0002$). The G-T haplotype was similarly associated in a set of never-smoking subjects from the Family Heart Study (FEV₁, $p = 0.03$; FVC, $p = 0.03$).

Conclusions: The replication of the association in two populations supports the possibility that *SMOC2* might play an important role in the determination of FEV₁ and FVC.

Keywords: FEV₁; FVC; genetics; single-nucleotide polymorphism

The observation that pulmonary diseases cluster in families (1, 2) and the substantial heritability of spirometric pulmonary function measures demonstrated in population-based samples (3–10) support the hypothesis that genetic factors influence both variability in pulmonary function and risk of chronic obstructive pulmonary disease (COPD). The spirometric measurements of FEV₁, FVC, and FEV₁/FVC are estimated to be between 15 and 60% heritable in population-based samples (8, 10). Familial factors may influence lung size during development, as well as affect the response to environmental toxins such as cigarette smoke.

Today, α_1 -antitrypsin deficiency is the only proven genetic determinant of COPD (11–16). The homozygous deficiency of the serine protease inhibitor α_1 -antitrypsin is associated with early-onset emphysema in smokers in their fourth decade of life and nonsmokers in their fifth decade (17) and accounts for less

AT A GLANCE COMMENTARY

Scientific Knowledge on the Subject

α_1 -Antitrypsin deficiency is the only proven genetic determinant of chronic obstructive pulmonary disease.

What This Study Adds to the Field

The study provides evidence for the implication of yet another gene, *SMOC2*, in lung function. *SMOC2* was found to affect FEV₁ and FVC in two different populations, the NHLBI Framingham Heart Study and the NHLBI Family Heart Study.

than 2% of all COPD (18). The heterozygous form of the mutation has been inconsistently associated with an increased risk of COPD. Mutations in alleles of α_1 -antichymotrypsin, a highly homologous protease inhibitor (19), have also been associated with obstructive lung disease in case-control studies (20).

Genomewide linkage for pulmonary function measures has been performed in population-based family studies. In the Framingham Heart Study (FHS), linkage was identified on chromosome 6qter: FEV₁ (LOD [logarithm of the odds] = 2.4) and FEV₁/FVC (LOD = 1.4) (21). A follow-up study demonstrated that the addition of new markers resulted in stronger evidence for linkage to FEV₁ at 184.5 cM (LOD = 5.0) (22). The linkage in FHS lies in the region of 184 cM (D6S503) to 190 cM (D6S281) on 6q27. Linkage in this region was not reported in the population-based sample from the National Heart, Lung, and Blood Institute (NHLBI) Family Heart Study (23). To identify the possible gene on 6q27 that is contributing to the linkage peak observed in the Framingham sample, we adopted a candidate gene approach. One candidate gene is the “secreted protein acidic and rich in cysteines” (SPARC)-related modular calcium binding 2 gene (*SMOC2*).

SMOC2 (locus ID: 64094) harbors a Kazal domain, two thymoglobulin type-1 domains, two EF-hand calcium-binding domains, and a putative signal peptide (24). The *SMOC2* Kazal domain, like α_1 -antitrypsin, encodes for a serine protease inhibitor. The thymoglobulin type-1 domains might also act as inhibitors of several proteases. The *SMOC2* gene has been cloned and characterized (25). The protein was reported to be expressed in the lung and in the aorta, and the mRNA has been reported to be up-regulated during neointima formation in a rat balloon injury model (24), which suggests that *SMOC2* may play an important role in lesion growth. Here, we examine the association between 20 single-nucleotide polymorphisms (SNPs) spanning 1,477 kb around and in *SMOC2* and spirometry measures in the FHS population. In addition, we report a haplotype analysis of the implicated SNPs evaluated in a sample from the NHLBI Family Heart Study.

(Received in original form January 25, 2006; accepted in final form January 4, 2007)

The Framingham Heart Study is supported by NIH/NHLBI N01-HC-25195. J.B.W. is supported by a Young Clinical Scientist award from the Flight Attendant Medical Research Institute.

Correspondence and requests for reprints should be addressed to Samer Karamohamed, Ph.D., The University of Chicago, Department of Human Genetics, 920 East 58th Street, Room 515, Chicago, IL 60611. E-mail: samer@uchicago.edu

Am J Respir Crit Care Med Vol 175, pp 554–560, 2007

Originally Published in Press as DOI: 10.1164/rccm.200601-1100C on January 4, 2007
Internet address: www.atsjournals.org

Some of the results of these studies have been previously reported in the form of an abstract and poster at the American Society of Human Genetics 2003 annual meeting (26).

METHODS

FHS Subjects and Analysis

This study examined unrelated subjects from the FHS offspring cohort, a sample of white Americans of predominantly western European descent. In 1971, FHS recruited the biologic offspring of the original participants and the spouses of these offspring. A cohort of unrelated offspring participants was sampled by selecting one member from each family, yielding a sample of 1,888 individuals. The spirometric methods used in the FHS have been previously described (21, 22, 27). Spirometric data were available on 1,734 of the unrelated subjects. The mean value of spirometry and covariates at two time points was used when available. Participants having only one examination with spirometry were included with data from a single exam.

Analyses were performed using a multiple linear regression model with FEV₁ or FVC as the dependent variable and a dominant modeling strategy. The models included as covariates age, sex, height, body mass index (BMI) (kg/m²), smoking status (never, former, or current), and pack-years. In addition, SNPs were analyzed within strata of never- and ever-smokers using the same covariates. Haplotype association, adjusted for covariates, was assessed using the program haplo.stats (28, 29) (Haplo.stats software is available at <http://cran.r-project.org/src/contrib/Descriptions/haplo.stats.html>).

Family Heart Study Subjects and Analysis

The Family Heart Study participants evaluated in this report comprise 225 white families being studied for linkage to BMI on chromosome 7 (30). Details of the original design have been described previously (31). SNP and haplotype association was evaluated using Family Based Association Tests implemented in the FBAT program (32, 33) (FBAT software is available at <http://www.biostat.harvard.edu/~fbat/default.html>). The analysis was performed in the full sample and stratified by smoking using phenotypic residuals for FEV₁ and FVC that were adjusted for age, age-squared, BMI, height, smoking status, and pack-years in separate regression models by sex (23). The p values reported are FBAT results testing the null hypothesis of no linkage and no association. Estimates of the SNP and haplotype effects on FEV₁ and FVC in liters were generated using regression models adjusting for the same covariates. Haplotypes were predicted in families using Merlin (34) (Merlin software is available at <http://www.sph.umich.edu/csg/abecasis/Merlin/download/>).

SNP Genotyping

We evaluated 42 SNPs selected from the SNP Consortium Database (www.ncbi.nlm.nih.gov/SNP) using assays based on the Sequenom matrix-assisted laser desorption/ionization time-of-flight mass spectrometry platform (35). All SNPs were genotyped in at least 32 individuals to establish their heterozygosity. Polymorphic SNPs were typed in all unrelated FHS subjects. Lab controls and duplicate samples were genotyped for quality control. All SNPs were tested for Hardy-Weinberg equilibrium (HWE). Two SNPs exhibiting modest departure from

HWE were regenotyped using TaqMan assays on a Prism 7900HT Sequence Detection System (Applied Biosystems, Foster City, CA). The ABI Prism system was also used for genotyping in the Family Heart Study subjects.

Linkage Disequilibrium and Haplotype Analysis

Linkage disequilibrium (LD) between all markers was assessed using Haploview (36) (Haploview software is available at <http://www.broad.mit.edu/mpg/haploview/download.php>). Haplotype blocks were estimated using the confidence interval method (37). The default settings were used to define SNP pairs in strong LD. A block was identified when at least 95% of SNP pairs in a region met criteria for strong LD. Haplotype block information was used to select SNPs for haplotype analysis. SNP rs747995 was selected for the haplotype because of the association observed in nonsmokers, and rs1402 was selected for its position outside of the LD block that included rs747995 and due to its observed association with phenotype in the complete sample. These two SNPs generated a strong haplotype result in the Framingham sample and were genotyped and tested for association in the Family Heart Study. To explore haplotypes among smokers, rs2248421, rs1402, and rs968440, each identified for association with FEV₁ in the ever-smokers and representing different LD blocks, were combined.

RESULTS

Characteristics of the unrelated FHS and the Family Heart Study subjects are presented in Tables 1 and 2. In both populations, nonsmoking subjects had a higher FEV₁/FVC ratio than smokers. SNPs in and near the *SMOC2* gene were selected from the National Center for Biotechnology Information (NCBI) database on the basis of their location. A total of 42 SNPs were genotyped; 15 SNPs were found to be nonpolymorphic in the initial screening test. The remaining 27 SNPs were typed in all FHS subjects; 7 were out of HWE ($p < 0.01$) and excluded from further analyses. SNPs rs747995 and rs714937 had moderate departure from HWE ($0.01 < p < 0.05$) and were retyped on the ABI platform. The genotypes between the Sequenom and ABI were largely consistent, with 98.5% identical genotypes for rs747995, and 99.1% identical genotypes for rs1402. The mismatched genotypes were discarded and only concordant calls were used in association analyses.

FHS Sample

Table 3 shows characteristics of the 20 SNPs in and around *SMOC2* that were used for association analysis. The SNPs are presented according to their chromosomal location in build 36.1. All tested SNPs in *SMOC2* are intronic and their minor allele frequency is between 3 and 44%. Only one SNP (rs1402) showed significant association to FEV₁ ($p = 0.003$) and FVC ($p = 0.02$) in the complete sample with the presence of the minor allele (G) associated with a mean FEV₁ 96 ml higher and mean FVC 84 ml higher than common allele homozygotes (Table 3). A second SNP, rs747995, was significantly associated with only

TABLE 1. CHARACTERISTICS OF THE UNRELATED NHLBI FRAMINGHAM SUBJECTS

	Nonsmokers		Smokers	
	Males (n = 261)	Females (n = 317)	Males (n = 608)	Females (n = 548)
Age, yr	51.4 ± 10.0	52.8 ± 9.9	53.6 ± 9.4	52.0 ± 9.1
Body mass index	27.5 ± 3.7	26.7 ± 5.5	28.1 ± 4.0	25.8 ± 4.9
FEV ₁ , % predicted	99 ± 0.14	101 ± 0.13	94 ± 0.15	96 ± 0.16
FVC, % predicted	100 ± 0.13	101 ± 0.13	97 ± 0.13	99 ± 0.13
FEV ₁ /FVC ratio	0.76 ± 0.07	0.77 ± 0.05	0.74 ± 0.07	0.75 ± 0.07
Current smoking, %			35.9	42.9
Pack-years			28.8 ± 22.7	20.3 ± 19.4

Values are mean ± SD, except % current smoking.

TABLE 2. CHARACTERISTICS OF THE NHLBI FAMILY HEART STUDY SUBJECTS

	Nonsmokers		Smokers	
	Males (n = 248)	Females (n = 363)	Males (n = 251)	Females (n = 194)
Age, yr	50.1 ± 14.4	52.7 ± 13.6	54.4 ± 13.3	52.5 ± 12.5
Body mass index	28.1 ± 5.0	28.3 ± 7.1	27.9 ± 4.7	27.8 ± 6.7
FEV ₁ , % predicted	101 ± 14	102 ± 15	89 ± 18	97 ± 19
FVC, % predicted	106 ± 14	108 ± 15	98 ± 15	106 ± 17
FEV ₁ /FVC ratio	0.76 ± 0.07	0.77 ± 0.06	0.72 ± 0.10	0.74 ± 0.08
Current smoking, %			31.5	29.4
Pack-years			33.1 ± 28.5	19.8 ± 19.0

Values are mean ± SD, except % current smoking.

FVC. After stratification by never/ever smoking status, however, seven SNPs showed significant association ($p < 0.05$) with FEV₁ or FVC in at least one of the strata (Table 4). In nonsmokers, SNP rs747995 showed the strongest association with FEV₁ ($p = 0.0006$) and FVC ($p = 0.0008$), with the minor allele associated with a 149 ml higher FEV₁ and 177 ml higher FVC compared with common allele homozygotes.

The region spanning the genotyped SNPs was assessed for patterns of LD, and three LD blocks were identified (Figure 1). Block 1 harbored four SNPs, including rs747995 (SNPs 6–9 in Figure 1), and the three SNPs in this block with minor allele frequencies (MAF) between 14 and 20% each showed association with FEV₁ and FVC in nonsmokers. The lower MAF SNP in this block (rs1884475, MAF = 5%) was modestly associated with FVC in smokers. SNP rs1402 was in the region adjacent to this block. A haplotype of SNPs rs747995 and rs1402 revealed a significant global haplotype association with FEV₁ ($p = 0.02$) and FVC ($p = 0.02$) using all unrelated FHS subjects and a 0.05 significance level. The haplotype association was stratified by

smoking status, and although the global haplotype association in ever-smokers did not suggest association (FEV₁, $p = 0.19$; FVC, $p = 0.59$), the global haplotype association in the strata of never-smokers was statistically significant (FEV₁, $p = 0.0008$; FVC, $p = 0.002$). Table 5 presents the different haplotypes of rs747995–rs1402 in never-smokers, the haplotype frequency, β -estimates, and p values for FEV₁ and FVC. In never-smoking subjects, one common two-SNP haplotype (77%), representing alleles G and T from SNPs rs747995 and rs1402, respectively, was found to be associated with 126 ml lower FEV₁ ($p = 0.0002$) and 157 ml lower FVC ($p = 0.0002$).

A haplotype incorporating the rs2248421, rs1402, and rs968440 SNPs was significantly associated with FEV₁ and FVC ($p = 0.006$ for both) in the stratum of smokers (Table 6). The three-SNP haplotype result identified a 1.7% frequent haplotype (T-T-A) associated with 255 ml lower mean FEV₁ and 294 ml lower mean FVC in smokers. This haplotype was not globally significant in the nonsmokers, and the effect of the T-T-A haplotype in nonsmokers was estimated to be only a 19 ml lower mean FEV₁.

TABLE 3. GENOTYPED SINGLE-NUCLEOTIDE POLYMORPHISMS ON 6q27 AND THEIR ASSOCIATION WITH PULMONARY FUNCTION MEASURES (FEV₁ AND FVC) IN NHLBI FRAMINGHAM HEART STUDY SUBJECTS

SNP Number	SNP rs Cluster ID	Chromosomal Location	Chromosomal Region	Minor Allele	Minor Allele Frequency (%)	FEV ₁		FVC	
						β - Estimate	p Value	β - Estimate	p Value
1	rs1548349	168202694	Coding <i>FRMD1</i>	G	0.083	0.023	0.49	0.006	0.88
2	rs902393	168206473	Coding <i>FRMD1</i>	A	0.086	0.023	0.42	0.029	0.38
3	rs1232314	168632154	Intron 1 <i>SMOC2</i>	G	0.216	-0.014	0.52	-0.016	0.52
4	rs2248785	168656395	Intron 2 <i>SMOC2</i>	A	0.045	-0.033	0.39	-0.048	0.28
5	rs2248421	168659271	Intron 2 <i>SMOC2</i>	T	0.037	-0.054	0.2	-0.095	0.05
6	rs2255680	168670298	Intron 3 <i>SMOC2</i>	A	0.197	0.009	0.7	0.024	0.36
7	rs747995	168673665	Intron 4 <i>SMOC2</i>	A	0.143	0.038	0.17	0.066	0.04
8	rs1884475	168673825	Intron 4 <i>SMOC2</i>	T	0.046	-0.024	0.54	-0.063	0.16
9	rs714937	168674344	Intron 4 <i>SMOC2</i>	A	0.195	0.005	0.84	0.019	0.48
10	rs1402	168686225	Intron 4 <i>SMOC2</i>	G	0.07	0.096	0.003	0.084	0.02
11	rs753149	168722764	Intron 7 <i>SMOC2</i>	T	0.049	-0.033	0.36	0.009	0.84
12	rs968440	168782529	Intron 9 <i>SMOC2</i>	G	0.339	0.029	0.19	0.043	0.09
13	rs911948	168783978	Intron 9 <i>SMOC2</i>	A	0.309	0.028	0.19	0.028	0.26
14	rs196450	168815325	3' <i>SMOC2</i>	C	0.429	-0.014	0.55	-0.005	0.85
15	rs196458	168859305	3' <i>SMOC2</i>	G	0.078	-0.012	0.69	0.002	0.94
16	rs196472	168867264	3' <i>SMOC2</i>	A	0.222	-0.034	0.13	-0.043	0.09
17	rs1407516	168869733	3' <i>SMOC2</i>	C	0.249	-0.003	0.9	-0.009	0.72
18	rs1401	169175653	3' <i>SMOC2</i>	T	0.303	-0.018	0.42	-0.02	0.42
19	rs736190	169522023	3' <i>SMOC2</i>	A	0.165	0.006	0.8	-0.005	0.86
20	rs1001844	169679886	Intron <i>WDR27</i>	T	0.058	-0.016	0.65	-0.03	0.47

Definition of abbreviation: SNP = single-nucleotide polymorphism.

All SNPs were identified in the public SNP database (www.ncbi.nlm.nih.gov/SNP) and are presented by their cluster identification numbers. The physical position (in base pairs) and chromosomal regions correspond to genome build 36.1. The association analysis for all subjects is presented. β -Estimates and p values for FEV₁ and FVC are presented. Unit β values are represented in liters with a negative sign associated with a lower mean FEV₁ and FVC in minor allele carriers. p Values ≤ 0.05 are shown in bold.

TABLE 4. SINGLE-NUCLEOTIDE POLYMORPHISM ASSOCIATION WITH FEV₁ AND FVC IN NEVER- AND EVER-SMOKING NHLBI FRAMINGHAM HEART STUDY SUBJECTS

SNP ID	Never-smoking				Ever-smoking				
	FEV ₁		FVC		FEV ₁		FVC		
	β- Estimate	p Value	β- Estimate	p Value	β- Estimate	p Value	β- Estimate	p Value	
1	rs1548349	-0.067	0.21	-0.032	0.62	0.069	0.10	0.025	0.59
2	rs902393	-0.008	0.86	-0.007	0.91	0.036	0.32	0.044	0.28
3	rs1232314	-0.03	0.41	-0.028	0.52	-0.007	0.81	-0.01	0.75
4	rs2248785	0.007	0.91	0.052	0.5	-0.053	0.26	-0.094	0.08
5	rs2248421	0.042	0.55	0.046	0.59	-0.102	0.05	-0.161	0.007
6	rs2255680	0.092	0.01	0.135	0.002	-0.037	0.2	-0.037	0.25
7	rs747995*	0.149	0.0006	0.177	0.0008	-0.022	0.53	0.006	0.89
8	rs1884475	0.057	0.39	0.075	0.35	-0.058	0.23	-0.122	0.02
9	rs714937*	0.076	0.04	0.122	0.006	-0.034	0.25	-0.039	0.25
10	rs1402	0.09	0.08	0.118	0.06	0.103	0.01	0.069	0.13
11	rs753149	-0.003	0.95	0.034	0.62	-0.046	0.32	-0.005	0.93
12	rs968440	-0.018	0.62	0.025	0.56	0.05	0.05	0.057	0.06
13	rs911948	0.019	0.59	-0.006	0.88	0.034	0.21	0.045	0.14
14	rs196450	-0.023	0.55	0.01	0.82	-0.011	0.7	-0.014	0.69
15	rs196458	-0.085	0.09	-0.036	0.55	0.024	0.52	0.02	0.64
16	rs196472	-0.056	0.11	-0.071	0.09	-0.023	0.4	-0.03	0.34
17	rs1407516	-0.009	0.8	-0.002	0.96	0.0003	0.99	-0.011	0.72
18	rs1401	-0.043	0.22	-0.026	0.54	-0.002	0.93	-0.015	0.62
19	rs736190	0.0009	0.98	0.024	0.6	0.008	0.79	-0.02	0.56
20	rs1001844	0.019	0.74	0.075	0.28	-0.031	0.49	-0.078	0.12

Definition of abbreviation: SNP = single-nucleotide polymorphism.

* SNPs genotyped using both Sequenom and ABI platforms and cleaned for any genotype inconsistencies between methods.

p Values ≤ 0.05 are shown in bold.

Family Heart Study Sample

Similar to the FHS, the Family Heart Study is a population-based family cohort, and the lung function characteristics shown in Table 2 show similarity to the Framingham sample. We attempted to replicate the SNP and haplotype association observed in never-smokers in a sample from the NHLBI Family Heart Study. SNPs rs747995 and rs1402 were tested using a dominant/recessive model in FBAT in the full sample of never- and ever-smokers, and the minor allele of rs1402 was associated with both FEV₁ (p = 0.05) and FVC (p = 0.03). The presence of the minor allele (G) of rs1402 was associated with a higher mean FEV₁ by 85 ml and FVC by 100 ml compared with the common allele homozygotes. In these samples, rs747995 was significantly associated with FVC (p = 0.02), with the minor allele (A) associated with higher FVC levels for the homozygous genotype. A total of 21 individuals carried the AA genotype of rs747995, and they had a mean FVC 264 ml higher than carriers of the G allele. In the full sample, the haplotype of rs747995 and rs1402 was borderline globally significant for association with FVC (p = 0.07). In never-smokers, the G-T haplotype of rs747995 and rs1402 showed association to FEV₁ and FVC (p = 0.03 for both), although the global haplotype tests did not produce significant results (FEV₁, p = 0.14; FVC, p = 0.06). Table 5 presents the different haplotype results for NHLBI Family Heart Study non-smoking subjects using an additive genetic model for comparison with FHS results. Similar to the observed effects in Framingham, the Family Heart study results indicate the common haplotype (80%) to be associated with lower mean levels of FEV₁ and FVC and a 7% haplotype associated with higher levels.

Bioinformatic analysis for the region harboring the haplotype showed that rs747995 and rs1402 reside in intron 4 of *SMOC2*. Figure 2 shows a genomic view of the two haplotyped SNPs and their location in *SMOC2*. Further analysis showed that the region of *SMOC2* intron 4 harbors an antisense encoding mRNA (AK130868) (Figure 2).

DISCUSSION

Using linkage analysis, we previously identified a quantitative trait locus influencing FEV₁ on 6q27 (LOD score = 5 at 184.5 cM) (22). In an effort to identify a gene(s) in the linkage region that influences FEV₁ and/or FVC, we applied a candidate gene approach. One candidate gene on 6q27 is *SMOC2*, which has homology to the protease inhibitor α₁-antitrypsin, the one gene known to influence the development of COPD in smokers.

The *SMOC2* homology to α₁-antitrypsin is characterized by the domain homology. *SMOC2* encodes an EF-hand calcium-binding domain, two thymoglobulin-like domains, a follistatin-like domain, and a novel domain without known homologs (25). The thymoglobulin domains bind and act as inhibitors to different proteases, such as serine and cysteine proteases (38, 39). The follistatin-like domains, including the Kazal-type protease inhibitor domain, are usually indicative of serine protease inhibitors (such as the thymoglobulin domain) (40). Thus, *SMOC2* contains multiple domains with the potential to act as protease inhibitors, and protease/antiprotease imbalances represent a leading hypothesis in the etiology of COPD.

To assess the role of *SMOC2* in association with FEV₁ and FVC, we studied a set of 1,734 unrelated individuals from the FHS (Table 1). Although a significant association with both FEV₁ and FVC was found in the total sample for only 1 of the 20 SNPs tested, an additional 6 SNPs showed significant association with lung function when the sample was stratified on smoking status. SNPs that showed significant association to FEV₁ and FVC in nonsmokers (rs2255680, rs747995, and rs714937) were located in introns 3 and 4 of *SMOC2*. SNP rs2248421, which resides in intron 2, showed significant association to FVC in ever-smokers, and rs1884475 and rs1402 (intron 4) were also associated in the ever-smoking sample. HapMap data suggest that 21 haplotype blocks may be present within *SMOC2* (<http://www.hapmap.org/>). Although we may not have sampled SNPs from each of these blocks to rule out other regions

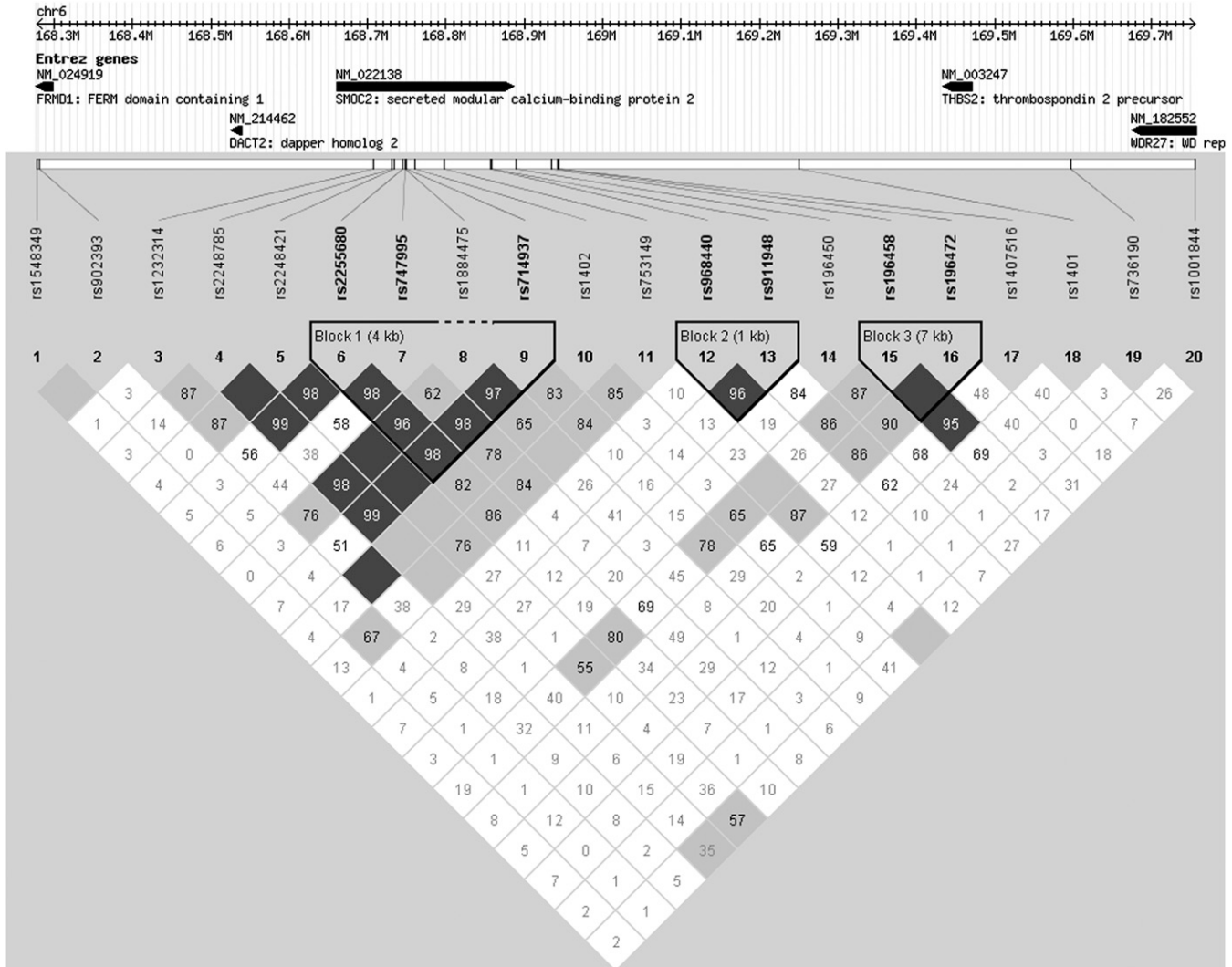


Figure 1. Linkage disequilibrium (LD) block structure around *SMOC2*. The color scheme (white to black) represents the increasing strength of LD. Values for D' ($\times 100$) are shown, and empty boxes indicate $D' = 1$. Three LD blocks were identified using the confidence interval method.

TABLE 5. HAPLOTYPE ASSOCIATION RESULTS FOR SINGLE-NUCLEOTIDE POLYMORPHISMS IN THE *SMOC2* GENE WITH FEV₁ AND FVC IN NEVER-SMOKERS FROM THE NHLBI FRAMINGHAM AND FAMILY HEART STUDY

rs747995	rs1402	Frequency	FEV ₁		FVC	
			β -Estimate	p Value	β -Estimate	p Value
Framingham Heart Study Unrelated Subjects						
G	T	0.77	-0.126	0.0002	-0.157	0.0002
G	G	0.07	0.092	0.10	0.139	0.04
A	T	0.16	0.124	0.002	0.140	0.004
Family Heart Study Subjects						
G	T	0.80	-0.062	0.03	-0.122	0.03
G	G	0.07	0.169	0.05	0.186	0.05
A	T	0.13	0.001	0.29	0.077	0.29

The different haplotypes for rs747995–rs1402 single-nucleotide polymorphisms, the haplotype frequency, and the p value for each haplotype with FEV₁ and FVC in the Framingham and Family Heart Study are presented. Haplotype-specific β -estimates with a negative sign are associated with lower mean FEV₁ (L) and FVC (L) and those without signs are associated with higher mean FEV₁ (L) and FVC (L). p Values ≤ 0.05 are shown in bold.

of the gene as harboring associated SNPs, the area of strongest association was localized to the region around exons 3 and 4. The *SMOC2* region was assessed for patterns of LD to identify SNPs for haplotype analysis (Figure 1). We identified a common

TABLE 6. HAPLOTYPE ASSOCIATION ANALYSIS FOR SINGLE-NUCLEOTIDE POLYMORPHISMS IN *SMOC2* WITH FEV₁ AND FVC IN EVER-SMOKERS FROM UNRELATED SUBJECTS FROM THE NHLBI FRAMINGHAM HEART STUDY

rs2248421	rs1402	rs968440	Frequency	FEV ₁		FVC	
				β -Estimate	p Value	β -Estimate	p Value
T	T	A	0.017	-0.255	0.006	-0.294	0.003
C	T	A	0.597	-0.048	0.02	-0.045	0.06
T	T	G	0.022	-0.032	0.67	-0.123	0.12
C	G	G	0.022	0.114	0.14	0.088	0.33
C	G	A	0.045	0.092	0.07	0.065	0.26
C	T	G	0.297	0.048	0.03	0.064	0.01

Haplotype-specific β -estimates with a negative sign are associated with lower mean FEV₁ (L) and FVC (L) and those without signs are associated with higher mean FEV₁ (L) and FVC (L). p Values ≤ 0.05 are shown in bold.

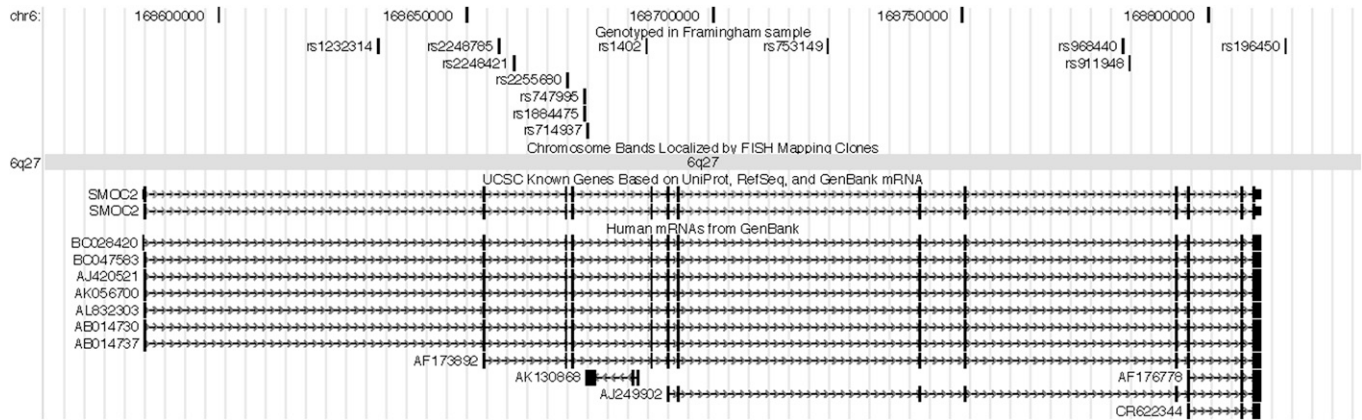


Figure 2. Genomic view of 6q27 harboring the *SMOC2* region. Single-nucleotide polymorphisms genotyped in the Framingham sample are labeled, the intron/exon structure of *SMOC2* is shown, and the *SMOC2* and AK130868 mRNA are depicted. (Available from the UCSC Genome Browser website: <http://genome.ucsc.edu>).

haplotype (rs747995–rs1402) with a frequency of 77% in the nonsmoking FHS subjects that was associated with lower FEV₁ and FVC levels ($p = 0.0002$ for both). Two minor haplotypes of 7 and 16% frequency were associated with higher FEV₁ and FVC levels.

To replicate these results in a separate sample, we genotyped the two SNPs in the associated haplotype (rs1402–rs747995) in a set of 225 pedigrees from the NHLBI Family Heart Study. Both SNPs individually were associated with pulmonary function in the full Family Heart Study sample. For rs1402, the genetic model and estimates of the effect on FEV₁ and FVC were consistent with the observations from the FHS. For rs747995, the minor allele association with higher levels of lung function was observed more clearly for carriers of two copies of the minor allele than one. This recessive model result compared with a dominant model result calls into question whether the heterozygote genotype has a replicated effect on phenotype, but in both studies, the minor allele homozygotes for rs747995 were observed with higher mean spirometry. Nonetheless, haplotype analysis of the two SNPs revealed that both the haplotype frequencies and their association with lung function among never-smokers were similar in the two samples. Haplotype association was observed despite the absence of linkage to pulmonary function on 6q27 in the Family Heart Study. The replication of the haplotype frequency and effect confirms that these SNPs are common in the general population and provides strong evidence that they are associated with FEV₁ and FVC.

The two SNPs we identified span a small region (~ 12,560 bp) on 6q27 that flank exon 4 of *SMOC2*. How this region impacts *SMOC2* gene function is not currently known. Effects on expression, splicing, or properties of the *SMOC2* gene product are possible. In addition, an identified antisense mRNA transcript (AK130868) in this region could impact readout from *SMOC2*. Further functional analyses using various combinations of these SNPs may help to confirm these findings and provide insight into the function of each variant and transcripts.

Because this was a population-based study with a low prevalence of overt obstructive lung disease, it was not possible to determine whether *SMOC2* haplotypes were associated with COPD *per se*. Different SNPs showed association to lung function in smokers and never-smokers, with the strongest association signal found in never-smoking subjects. Moreover, the effect of the alleles on lung function appears to be modified by exposure to cigarette smoke. We observed minor allele effect estimates

associated with higher lung function in never-smokers and lower lung function in smokers. If the haplotypes identified correspond to different *SMOC2* isoforms, those isoforms may influence lung function differently in the presence or absence of smoking. The association of the A-T and G-G haplotypes with higher lung function in nonsmokers may reflect a gene variant(s) with an impact on lung growth and development, whereas the association of haplotypes with reduced lung functions in smokers, if confirmed, may reflect a gene variant with an increased susceptibility to proteolytic lung injury. Whether *SMOC2* variants directly affect susceptibility to development of COPD either by modifying lung volumes attained during growth and development or by influencing response to environmental irritants such as tobacco smoke would be better addressed in a clinical sample.

One limitation of this study is the ability to address how the *SMOC2* variants explain the linkage result previously described. Due to scarce DNA resources, only a subset of the participants previously studied for linkage are available for new genotyping. Although we previously obtained an LOD score of 5 on 6q27, the smaller sample currently available generates only an LOD of 1.38. Adjusting for the rs747995 and rs1402 SNPs in linkage increased the LOD to 1.47, suggesting that although the SNPs account for variance in FEV₁, they do not necessarily account for the linkage observed.

In summary, the strong association of SNPs and haplotypes in the intron-4 region of *SMOC2* with the pulmonary function measurements FEV₁ and FVC suggests that common *SMOC2* sequence variants are associated with pulmonary function in the general population. This inference is strengthened by the identification of a common haplotype associated with lung function in two populations, unrelated FHS and Family Heart Study sample. Identification of the specific genetic variant affecting lung function may permit improved COPD risk assessment.

Conflict of Interest Statement: None of the authors has a financial relationship with a commercial entity that has an interest in the subject of this manuscript.

Acknowledgment: The authors thank the investigators and participants of the NHLBI Framingham and Family Heart Studies for making this work possible. The authors thank Dr. Richard H. Myers for facilitating the collaboration with the Family Heart Study and providing genotyping services.

References

1. 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.
2. Lomas DA, Silverman EK. The genetics of chronic obstructive pulmonary disease. *Respir Res* 2001;2:20-26.
 3. Hubert HB, Fabsitz RR, Feinleib M, Gwinn C. Genetic and environmental influences on pulmonary function in adult twins. *Am Rev Respir Dis* 1982;125:409-415.
 4. Lewitter FI, Tager IB, McGue M, Tishler PV, Speizer FE. Genetic and environmental determinants of level of pulmonary function. *Am J Epidemiol* 1984;120:518-530.
 5. McClearn GE, Svartengren M, Pedersen NL, Heller DA, Plomin R. Genetic and environmental influences on pulmonary function in aging Swedish twins. *J Gerontol* 1994;49:264-268.
 6. Higgins M, Keller J. Familial occurrence of chronic respiratory disease and familial resemblance in ventilatory capacity. *J Chronic Dis* 1975;28:239-251.
 7. Devor EJ, Crawford MH. Family resemblance for normal pulmonary function. *Ann Hum Biol* 1984;11:439-448.
 8. Ghio AJ, Crapo RO, Elliott CG, Adams TD, Hunt SC, Jensen RL, Fisher AG, Afman GH. Heritability estimates of pulmonary function. *Chest* 1989;96:743-746.
 9. Givelber RJ, Couropmitree NN, Gottlieb DJ, Evans JC, Levy D, Myers RH, O'Connor GT. Segregation analysis of pulmonary function among families in the Framingham study. *Am J Respir Crit Care Med* 1998;157:1445-1451.
 10. Wilk JB, Djousse L, Arnett DK, Rich SS, Province MA, Hunt SC, Crapo RO, Higgins M, Myers RH. Evidence for major genes influencing pulmonary function in the NHLBI Family Heart Study. *Genet Epidemiol* 2000;19:81-94.
 11. Poller W, Meisen C, Olek K. DNA polymorphisms of the alpha 1-antitrypsin gene region in patients with chronic obstructive pulmonary disease. *Eur J Clin Invest* 1990;20:1-7.
 12. Tarjan E, Magyar P, Vaczi Z, Lantos A, Vaszar L. Longitudinal lung function study in heterozygous PiMZ phenotype subjects. *Eur Respir J* 1994;7:2199-2204.
 13. Seersholm N, Kok-Jensen A, Dirksen A. Decline in FEV₁ among patients with severe hereditary α_1 -antitrypsin deficiency type PiZ. *Am J Respir Crit Care Med* 1995;152:1922-1925.
 14. Seersholm N, Wencker M, Banik N, Viskum K, Dirksen A, Kok-Jensen A, Konietzko N. Does alpha1-antitrypsin augmentation therapy slow the annual decline in FEV₁ in patients with severe hereditary alpha1-antitrypsin deficiency? Wissenschaftliche Arbeitsgemeinschaft zur Therapie von Lungenerkrankungen (WATL) alpha1-AT Study Group. *Eur Respir J* 1997;10:2260-2263.
 15. Piitulainen E, Eriksson S. Decline in FEV₁ related to smoking status in individuals with severe alpha1-antitrypsin deficiency (PiZZ). *Eur Respir J* 1999;13:247-251.
 16. Burdon JGW, Brenton S, Hocking V, Knight KR, Ayad M, Cook L, Jans ED. Decline in FEV₁ in patients with PiZ alpha1-antitrypsin deficiency: the Australian experience. *Respirology* 2002;7:51-55.
 17. Tobin MJ, Cook PJ, Hutchison DC. Alpha 1-antitrypsin deficiency: the clinical and physiological features of pulmonary emphysema in subjects homozygous for Pi type Z. A survey by the British Thoracic Association. *Br J Dis Chest* 1983;77:14-27.
 18. Snider GL. Molecular epidemiology: a key to better understanding of chronic obstructive lung disease. *Monaldi Arch Chest Dis* 1995;50:3-6.
 19. Chandra T, Stackhouse R, Kidd VJ, Robson KJ, Woo SL. Sequence homology between human alpha 1-antichymotrypsin, alpha 1-antitrypsin, and antithrombin III. *Biochemistry* 1993;22:5055-5061.
 20. Lindmark BE, Arborelius MJ, Eriksson SG. Pulmonary function in middle-aged women with heterozygous deficiency of the serine protease inhibitor alpha 1-antichymotrypsin. *Am Rev Respir Dis* 1990;141:884-888.
 21. Joost O, Wilk JB, Cupples LA, Harmon M, Shearman AM, Baldwin CT, O'Connor GT, Myers RH, Gottlieb DJ. Genetic loci influencing lung function: a genome-wide scan in the Framingham study. *Am J Respir Crit Care Med* 2002;165:795-799.
 22. Wilk JB, DeStefano AL, Joost O, Myers RH, Cupples LA, Slater K, Atwood LD, Heard-Costa NL, Herbert A, O'Connor GT, et al. Linkage and association with pulmonary function measures on chromosome 6q27 in the Framingham Heart Study. *Hum Mol Genet* 2003;12:2745-2751.
 23. Wilk JB, DeStefano AL, Arnett DK, Rich SS, Djousse L, Crapo RO, Leppert MF, Province MA, Cupples LA, Gottlieb DJ, et al. Genome-wide scan of pulmonary function measures in the National Heart, Lung, and Blood Institute Family Heart Study. *Am J Respir Crit Care Med* 2003;167:1528-1533.
 24. Nishimoto S, Hamajima Y, Toda Y, Toyoda H, Kitamura K, Komurasaki T. Identification of a novel smooth muscle associated protein, smap2, upregulated during neointima formation in a rat carotid endarterectomy model. *Biochim Biophys Acta* 2002;1576:225-230.
 25. Vannahme C, Gosling S, Paulsson M, Maurer P, Hartmann U. Characterization of SMOC-2, a modular extracellular calcium-binding protein. *Biochem J* 2003;373:805-814.
 26. Karamohamed S, Wilk JB, Shoemaker CM, Liu J, Myers RH, Gottlieb DJ. Linkage and association of polymorphisms in the SMOC2 gene with pulmonary function: the Framingham Heart Study [abstract]. *Am J Hum Genet* 2003;73:483.
 27. Sorlie PD, Kannel WB, O'Connor G. Mortality associated with respiratory function and symptoms in advanced age: the Framingham study. *Am Rev Respir Dis* 1989;140:379-384.
 28. Schaid DJ. Relative efficiency of ambiguous vs. directly measured haplotype frequencies. *Genet Epidemiol* 2002;23:426-443.
 29. 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.
 30. Jiang Y, Wilk JB, Borecki I, Williamson S, DeStefano AL, Xu G, Liu J, Ellison RC, Province M, Myers RH. Common variants in the 5' region of the leptin gene are associated with body mass index in men from the NHLBI Family Heart Study. *Am J Hum Genet* 2004;75:220-230.
 31. Higgins M, Province M, Heiss G, Eckfeldt J, Ellison RC, Folsom AR, Rao DC, Sprafka JM, Williams R. NHLBI Family Heart Study: objectives and design. *Am J Epidemiol* 1996;143:1219-1228.
 32. Horvath S, Wei E, Xu X, Palmer LJ, Baur M. Family-based association test method: age of onset traits and covariates. *Genet Epidemiol* 2001;21:S403-S408.
 33. Horvath S, Xu X, Lake SL, Silverman EK, Weiss ST, Laird NM. Family-based tests for associating haplotypes with general phenotype data: application to asthma genetics. *Genet Epidemiol* 2004;26:61-69.
 34. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin: rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 2002;30:97-101.
 35. Buetow KH, Edmonson M, MacDonald R, Clifford R, Yip P, Kelley J, Little DP, Strausberg R, Koester H, Cantor CR, et al. High-throughput development and characterization of a genomewide collection of gene-based single nucleotide polymorphism markers by chip-based matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Proc Natl Acad Sci USA* 2001;98:581-584.
 36. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005;21:263-265.
 37. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, et al. The structure of haplotype blocks in the human genome. *Science* 2002;296:2225-2229.
 38. Ritonja A, Turk B, Turk V. Equistatin, a new inhibitor of cysteine proteinases from *Actinia equina*, is structurally related to thyroglobulin type-1 domain. *J Biol Chem* 1997;272:13899-13903.
 39. Galesa K, Pain R, Jongsma MA, Turk V, Lenarcic B. Structural characterization of thyroglobulin type-1 domains of equistatin. *FEBS Lett* 2003;539:120-124.
 40. Williamson MP, Marion D, Wuthrich K. Secondary structure in the solution conformation of the proteinase inhibitor IIA from bull seminal plasma by nuclear magnetic resonance. *J Mol Biol* 1984;173:341-359.