Cryptic haplotypes of SERPIN A1 confer susceptibility to chronic obstructive pulmonary disease.

HUMAN MUTATION 27(1), 103–109, 2006

RESEARCH ARTICLE

Cryptic Haplotypes of *SERPINA1* Confer Susceptibility to Chronic Obstructive Pulmonary Disease

Sally Chappell,[1] Leslie Daly,[2] Kevin Morgan,[1] Tamar Guetta Baranes,[1] Josep Roca,[4] Roberto Rabinovich,[4] Ann Millar,[5] Seamas C. Donnelly,[3] Vera Keatings,[6] William MacNee,[7] Jan Stolk,[8] Pieter Hiemstra,[8] Massimo Miniati,[9] Simonetta Monti,[9] Clare M. O'Connor,[3] and Noor Kalsheker[1]*

[1]Division of Clinical Chemistry, Molecular Medical Sciences, Institute of Genetics, University Hospital, Queens Medical Centre, University of Nottingham, Nottingham, United Kingdom; [2]Department of Public Health Medicine and Epidemiology, Conway Institute, University College Dublin, Dublin, Ireland; [3]Department of Medicine and Therapeutics, Conway Institute, University College Dublin, Dublin, Ireland; [4]Service de Pneumologia, Hospital Clinic, Hospital Clinico y Provincial de Barcelona, Barcelona, Spain; [5]Lung Research Group, Department of Clinical Science at North Bristol, Southmead Hospital, University of Bristol, Bristol, United Kingdom; [6]Letterkenny General Hospital, Letterkenny, Ireland; [7]Respiratory Medicine, ELEGI Colt Laboratories, University of Edinburgh, Edinburgh, United Kingdom; [8]Department of Pulmonology (C3-P), Leiden University Medical Center, Leiden, The Netherlands; [9]CNR Institute of Clinical Physiology, Pisa, Italy

Communicated by Michael Dean

Chronic obstructive pulmonary disease (COPD) is a major cause of mortality and morbidity worldwide. While cigarette smoking is a major cause of COPD, only 15% of smokers develop the disease, indicating major genetic influences. The most widely recognized candidate gene in COPD is *SERPINA1*, although it has been suggested that *SERPINA3* may also play a role. To detect cryptic genetic variants that might contribute to disease, we identified 15 SNP haplotype tags from high-density SNP maps of the two genes and evaluated these SNPs in the largest case-control genetic study of COPD conducted so far. For *SERPINA1*, six newly identified haplotypes with a common backbone of five SNPs were found to increase the risk of disease by six- to 50-fold, the highest risk of COPD reported to date. In contrast, no haplotype associations for *SERPINA3* were identified. Hum Mutat 27(1), 103–109, 2006.

KEY WORDS: α-1-antitrypsin; SERPINA1; α-1-antichymotrypsin; SERPINA3; COPD; SNP; polymorphism; haplotype

INTRODUCTION

Chronic obstructive pulmonary disease (COPD [MIM# 606963]) is the sixth leading cause of death in the Western world [Gulsvik, 1999]. Although cigarette smoking is the major cause of COPD, only 15% of smokers develop the disease, which indicates that there may be genetic influences [Fletcher and Peto, 1977]. Protease–antiprotease imbalance is a widely supported hypothesis for the pathogenesis of COPD. The major circulating antiproteases, which have arisen by gene duplication, are SERPINs A1 [MIM# 107400] and A3 [MIM# 107280] [Brantly et al., 1988]. Severe deficiency of SERPINA1 accounts for about 2% of patients with COPD with a history of cigarette smoking, but the role of SERPINA3 remains to be clarified [Mittman et al., 1973]. Based on family studies, the estimated relative risk for genetic susceptibility to COPD is about 3 [DeMeo and Silverman, 2003]. Deficiency of SERPINA1 (more commonly known as α-1-antitrypsin) is the only recognized genetic factor contributing to disease. This manifests mainly in cigarette smokers, emphasizing the importance of genetic and environmental factors. The overall risk for COPD in cigarette smokers is increased approximately up to 20-fold in protease inhibitor (Pi) Z homozygotes of the *SERPINA1* gene, which occurs in about 1/2,000–3,000 individuals of northern European descent. The highest frequencies are observed in Scandinavia; however, not all Pi Z homozygotes develop COPD [Silverman et al., 1998]. A recent meta-analysis concluded that carriers of the Z variant have an increased odds ratio of 2.31 for COPD [Hersh et al., 2004].

The association of SERPINA1 deficiency with COPD has led to the protease–antiprotease imbalance theory of disease causation, which is supported by a number of experimental models [Karlinsky and Snider, 1978; Cavarra et al., 1996]. Briefly, this states that when neutrophils are recruited to the lung during an inflammatory process, they are primed to release the serine proteases neutrophil elastase (NE) and cathepsin G (CG). In animal models these proteases have been shown to be essential for killing bacteria

The Supplementary Material referred to in this article can be accessed at www.interscience.wiley.com/jpages/1059-7794/suppmat.

Received 9 June 2005; accepted revised manuscript 6 September 2005.

*Correspondence to: Professor Noor Kalsheker, Division of Clinical Chemistry, University Hospital, Nottingham, NG7 2UH, United Kingdom. E-mail: noor.kalsheker@nottingham.ac.uk

Grant sponsor: European Union 5th Framework Programme; Grant number: QLG1-CT-2001-01012.

Clare M. O'Connor and Noor Kalsheker are joint senior authors.

DOI 10.1002/humu.20275
Published online 8 November 2005 in Wiley InterScience (www.interscience.wiley.com).

[Reeves et al., 2002]. If sufficient inhibitor is not present, uninhibited enzyme activity can destroy the connective tissue of the lung, resulting in loss of normal tissue architecture and function. Given the role of this gene in COPD, there is a high prior probability of cryptic variation in the *SERPINA1* gene contributing to this disease. Genetic studies of *SERPINA1* variation other than the Z variant have yielded controversial results, which have been attributed to small sample sizes and a lack of adequate coverage of the gene [Poller et al., 1990; Sandford et al., 1997]. Similarly, a number of small case-control studies reported inconsistent results on the role of the other major circulating antiproteinase, SERPINA3 (more commonly known as α-1-antichymotrypsin), in COPD [Poller et al., 1992; Sandford et al., 1998].

In the largest case-control study reported to date, we screened 1,018 COPD patients and 911 nondiseased control smokers recruited at six European centers for nine SNPs that define the common haplotypes for the *SERPINA1* gene, plus the Pi S variant, which is the most common known mutation that causes mild deficiency in Europeans, and five SNPs that define common haplotypes of the *SERPINA3* gene.

MATERIALS AND METHODS

Subjects

The numbers of cases and controls recruited from each center were as follows: Barcelona, 70 controls and 138 cases; Bristol, 151 controls and 129 cases; Dublin, 195 controls and 196 cases; Edinburgh, 81 controls and 169 cases; Leiden, 216 controls and 188 cases; and Pisa, 198 controls and 198 cases. Approval for the study was obtained from the appropriate committees at each recruitment center (el Comite de Investigacion del Hospital Clinic, Barcelona; Southmead Local Research Ethics Committee, Bristol; St. Vincent's Hospital Ethics and Medical Research Committee, Dublin; Lothian Research Ethics Committee, Edinburgh; de Commissie Medische Ethiek van het LUMC, Leiden; and il Comitato Etico, Azienda Ospedaliera Pisana, Pisa). Informed consent was obtained from all subjects.

The criteria for patient recruitment were a firm clinical diagnosis of stable COPD, airflow limitation as indicated by $FEV_1 \leq 70\%$ normal predicted values (calculated as recommended [Crapo et al., 1981] and $FEV_1/FVC < 70\%$, no significant reversibility on bronchodilation [Brand et al., 1992], and a smoking history of ≥20 pack-years. Patients were excluded from the study if they had an established diagnosis of asthma, established obstructive syndrome or lung cancer, a history of atopy, known SERPINA1 deficiency, or a serum *SERPINA1* level of <1.0 g/L. They were also excluded if they had experienced an acute exacerbation in the 4 weeks preceding assessment for the study. Only white Caucasian patients were recruited.

Control subjects were recruited at each center to match COPD patients for age and gender. All controls were current or ex-smokers with a smoking history similar to that of the COPD patients. Individuals were excluded if they had a history of chronic lung disease or atopy, an acute pulmonary infection in the 4 weeks preceding assessment for the study, known SERPINA1 deficiency, a serum SERPINA1 level of <1.0 g/L, or a family history of COPD. Only individuals with no evidence of airflow obstruction (FEV_1 and $FVC \geq 80\%$, and $FEV_1/FVC > 70\%$) were included in the control group. Complete matching of control and COPD groups was not achieved because a high proportion of smokers over the age of 65 had evidence of some obstruction in pulmonary function. Even though these individuals did not have overt symptoms of disease, our strict criteria meant they were excluded from the study, which reduced the number of available controls.

The characteristics of each group were as follows, with values expressed as the mean (SD) when appropriate:

Controls: 63.5% male; age 60.8 (8.9); smoking pack-years 38.6 (17.4); %pred FEV1 95.3 (10.9); FEV/FVC% 77.9 (4.9).

COPD cases: 69.6% male; age 65.8 (8.2); smoking pack-years 48.9 (23.6); %pred FEV1 43.0 (15.3); FEV/FVC% 47.5 (12.2).

Identifying Haplotype-Tagging SNPs and Genotyping the Study Population

We previously identified nine haplotype-tagging SNPs that define common haplotypes for *SERPINA1* via high-density SNP mapping [Chappell et al., 2004]. In the current study we also screened for the Pi S variant (SNP 7) because even though it does not occur at a frequency of >5%, it is the most common known deficiency variant in Europeans.

Haplotype-tagging SNPs for the *SERPINA3* gene were identified using a strategy similar to that used for *SERPINA1*. Briefly, SNPs were identified and confirmed by bidirectional sequencing of all exons and 5′ flanking regions in 44 samples. SNPs that occurred at more than 5% were genotyped in 291 samples by the MRC Geneservice (Babraham, UK) using PCR with fluorescently-labeled Taqman probes (Vic or Fam labels).

Genotyping of the study population for the *SERPINA1* and *SERPINA3* haplotype-tagging SNPs was also carried out at the MRC Geneservice (Babraham, UK) using PCR with fluorescently-labeled Taqman probes (Vic or Fam labels). The primer and probe sequences are available on request. As a quality control measure, 44 samples of known genotype determined by sequencing were included. These genotypes were unknown to the MRC but were known at the source. When discrepancies were noted the analysis was repeated. This resulted in 100% concordance for all the assays.

The SNPs are numbered with respect to the GenBank genomic DNA sequences AL132708.3 for *SERPINA1*, and AL049839.3 for *SERPINA3*.

Analysis of Genetic Variation in Population Groups

Each of the SNPs in *SERPINA1* and *-A3* were analyzed for Hardy-Weinberg equilibrium (HWE) using PROC ALLELE in SAS/Genetics Release 8.2 (SAS Institute Inc., Cary, NC) with hot fix 82GN03. To examine linkage disequilibrium, the correlation coefficient between SNP pairs within each gene in cases and controls was calculated under the assumption of HWE using the same program.

Allele and Genotype Frequencies in COPD Cases and Controls

Analysis was performed using PROC CASECONTROL in SAS/Genetics. Since matching was not completely achieved on recruitment, we adjusted for any residual confounding due to age, sex, smoking, and center using logistic regression. This also gave us an increase in power by accounting for the matched design. Age was categorized into four groups (<55 years; 55–64 years; 65–69 years; 70+ years), and smoking was categorized as heavy or light based on a definition of heavy smoking that was center-dependent at an approximate 40th centile cut-point. All effects were coded as dummy binary variables, and interactions of age, sex, and smoking with center were included. This approach was also used to correct haplotype-disease associations (see below).

Haplotype Analysis

For a full analysis of the relationships between COPD and the *SERPINA1* and *-A3* SNPs, we used a staged, haplotype-based approach. It was first necessary to identify which SNP groupings within each gene should be subjected to the haplotype analysis. There are a number of approaches for detecting haplotype groups in a gene, but all are based on defining a group as a block of contiguous SNPs [Zhang and Jin, 2003]. The objective of these methods is to identify haplotype-tagging SNPs that can be used to reduce the number of SNPs to genotype while still achieving a large percentage coverage of variation within the gene. We were interested in identifying haplotypes that differed in frequency between COPD cases and controls, using haplotype-tagging SNPs already identified from high-density mapping of the *SERPINA1* and *-A3* genes. Since there is no reason to suspect that disease-associated haplotypes might be combinations of contiguous SNPs, and there are no proven shortcut methods for identifying such groupings, we screened all possible SNP groups in each gene for differences between cases and controls. This was done by performing an omnibus test [Zhao et al., 2000] as implemented in SAS/Genetics PROC HAPLOTYPE. Exact P-values based on a permutation test were employed. We used a screening value of 1,000 permutations to initially choose SNP groups with exact P-values ≤ 0.017. This ensured that (based on 0.0099 being the exact 95% lower binomial confidence limits for 17 out of 1,000) a nominal P-value of <0.01 was achieved. The P-value was then recalculated for these groups with 10,000 permutations, and the final groups were chosen based on a P-value of <0.01.

Once SNP groups that were potentially related to COPD were identified, employing the standard EM algorithm, PROC HAPLOTYPE was used to generate all possible haplotype patterns within these groups and to compare the distributions between cases and controls. Exact P-values based on 10,000 permutations were generated.

Haplotypes with a frequency of over 0.5% in cases and controls combined that differed between cases and controls at the $p<0.001$ level were considered relevant. This ensured that extremely rare haplotypes were excluded from consideration.

Confounder Adjustment

Because SNP data are unphased, it is not possible to assign haplotypes to individuals. PROC HAPLOTYPE uses the expectation maximization (EM) algorithm to determine the frequency of each possible haplotype in cases and controls, and performs an overall significance test [Zhao et al., 2000]. Without haplotype data at an individual level, however, only a limited confounder adjustment can be performed using a stratified approach. A recently described technique, termed haplotype trend regression (HTR), utilizes regression methods for confounder adjustment by assigning the probability of having a particular haplotype to each individual and substituting that probability for the usual 1/0 coded binary predictor variable of haplotype present or absent [Zaykin et al., 2002]. However, applying this technique to our data yielded odds ratios that were severely biased upwards, so we used an alternative unbiased approach. This approach uses a weighted logistic regression with the probability of having a particular haplotype defining the weights (see Supplementary Appendix S1; available online at www.interscience.wiley.com/jpages/1059-7794/suppmat). The haplotype probabilities were obtained using PROC HAPLOTYPE and used to adjust haplotype odds ratios and significance levels for the effects of center, age, sex, and smoking. Odds ratios (adjusted and nonadjusted) were based on the logistic regression and are the odds of disease in those with a particular haplotype relative to the odds in those without that haplotype.

Assessment of Population Stratification

To assess the effects of stratification we used data obtained from screening the same study population for 20 SNPs in four unlinked genes: γ-glutamyl cysteine synthetase catalytic and regulatory subunits, tumor necrosis factor, and epoxide hydrolase. Allelic association was first tested for each of the 20 SNPs using 2×2 contingency tables. The individual chi-square test statistics were summed for the 20 SNPs to give an overall summary chi-square statistic [Pritchard and Rosenberg, 1999].

Multiple Testing

Given the role of the *SERPINA1* gene in COPD, the prior probability of cryptic genetic variation contributing to disease would be considered to be high, and as a result the issue of multiple testing is not as problematic as it is in genomewide scans, where no a priori assumptions are made. We did not explicitly account for multiple testing in the calculation of P-values in this study; rather, we chose to use a stringent P-value and provide details of the analysis.

RESULTS

Identification of Haplotype-Tagging SNPs for *SERPINA3*

High-density SNP mapping of the SERPINA3 gene identified five SNPs that occurred with a frequency of >5%. These SNPs produce six major haplotypes representing almost 100% coverage of the variation observed in the gene (Table 1).

Allele and Genotype Frequencies in COPD and Control Populations

The locations of the 10 *SERPINA1* and five *SERPINA3* SNPs screened for in the study population are indicated in Figure 1. The allele frequencies for the minor variant of each of the 10 *SERPINA1* and five *SERPINA3* SNPs in the control and COPD groups are shown in Table 2. For ease of presentation, the SNPs are numbered sequentially relative to the coordinates of the human genomic DNA sequences AL132708.3 for *SERPINA1* and AL049839.3 for *SERPINA3*. The position and base change of each SNP relative to the hepatocyte transcription start site are also indicated in parentheses. SNP g.135728A>T (SNP 7) of *SERPINA1* is the Pi S variant.

The initial comparison of allele frequencies indicated that *SERPINA1* SNPs 1 and 7 differed significantly between controls and COPD groups. However, when corrected for smoking, age, gender, and center, only SNP 1 retained borderline significance (Table 2).

There were no differences between COPD and control groups in the frequency of any of the five *SERPINA3* SNPs before or after confounder adjustment (Table 2).

Only minor differences were detected in the frequencies of SNPs across the six recruitment sites for all 15 SNPs, with the exception of *SERPINA1* SNP 10, which varied from 4.85% in Edinburgh to 14.18% in Pisa. Previously published data showed population differences for this SNP, with a higher frequency in eastern Europe [Samilchuk et al., 1997]. There were, however, no significant differences in the frequency of this particular SNP

TABLE 1. Haplotypes of the *SERPINA3* Gene in Normal Subjects*

g.44695G>T (−12799)	g.57443G>T (−51)	g.59569G>A (2076)	g.59847A>G (2354)	g.64231C>G (6738)	Frequency
G	G	G	A	C	0.16
G	G	G	A	G	0.25
G	T	G	A	C	0.04
G	T	A	A	C	0.38
G	T	A	A	G	0.09
T	G	G	G	G	0.08
Rare Haplotypes (<1%)					<0.01

*Numbering is with respect to the GenBank sequence AL049839.3, with numbering with respect to the hepatocyte transcription start site shown in parentheses.

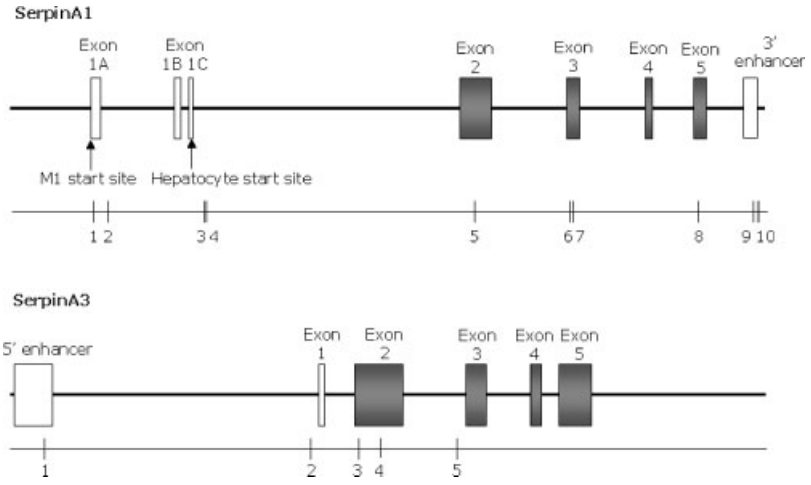

FIGURE 1. Location of SNPs in *SERPINA1* and *-A3*. Shaded boxes represent coding exons (not to scale).

TABLE 2. Allele Frequency of *SERPINA1* and *SERPINA3* SNP Variants in Controls and COPD Subjects*

SNP No	SNP	dbSNP reference	Control	COPD	P-value[a]	Corrected P-value[b]
SERPINA1						
1	g.126076T>C (−1973)	rs8004738	52.2%	48.7%	0.037	0.0294
2	g.126333C>T (−1716)	–	41.4%	44.5%	0.062	0.0528
3	g.128113C>T (65)	rs1243160	18.8%	18.5%	0.774	0.5455
4	g.128117G>A (69)	rs2854254	37.2%	35.1%	0.188	0.3619
5	g.133789G>A (5741)	rs709932	17.0%	17.0%	0.976	0.8276
6	g.135575T>C (7527)	rs6647	17.9%	19.6%	0.200	0.1880
7	g.135728A>T (7680)	rs17580	3.6%	5.0%	0.035	0.1825
8	g.138147A>C (9995)	–	26.5%	25.7%	0.576	0.4193
9	g.139425A>G (11377)	rs11832	48.2%	47.4%	0.643	0.7070
10	g.139535G>A (11487)	–	7.7%	7.8%	0.948	0.8418
SERPINA3						
1	g.44695G>T (−12799)	–	11.5	11.3	0.836	0.2763
2	g.57443G>T (−51)	rs1884082	52.1	50.5	0.324	0.7589
3	g.59569G>A (2076)	rs4934	49.8	47.3	0.121	0.3991
4	g.59847A>G (2354)	–	12.1	12.2	0.896	0.4972
5	g.64231C>G (6738)	rs8004988	44.5	43.0	0.353	0.1138

*Numbering is with respect to the GenBank sequence AL132708.3 for *SERPINA1* and AL049839.3 for *SERPINA3*, with numbering with respect to the hepatocyte transcription start site shown in parentheses.
[a]Crude P-values.
[b]P-values corrected by logistic regression for age, gender, center, and smoking.

between the cases (14.18%) and controls (13.26%) in the Pisa group or for any other population group. All of the SNPs reported were in HWE in both cases and controls.

The linkage disequilibrium (LD) results for *SERPINA1* in the control group are shown in Table 3, confirming in this large population the low values of LD that were previously noted in a smaller group for *SERPINA1* [Chappell et al., 2004]. The strength of LD between SNPs in *SERPINA3* is high in comparison to *SERPINA1* (Table 3).

Haplotype Analysis

For *SERPINA1* the initial analysis identified seven SNP groups that differed significantly between COPD cases and controls

TABLE 3. Linkage Disequilibrium Correlation Coefficients for *SERPINA1* and *SERPINA3* in Controls

SNP Nos.	2	3	4	5	6	7	8	9	10
SERPINA1									
1	0.855	0.156	0.188	0.211	0.206	0.118	0.158	0.356	0.103
2	–	0.247	0.094	0.271	0.047	0.096	0.229	0.247	0.136
3	–	–	0.621	0.198	0.026	0.095	0.234	0.313	0.281
4	–	–	–	0.236	0.067	0.151	0.083	0.046	0.107
5	–	–	–	–	0.205	0.075	0.660	0.354	0.131
6	–	–	–	–	–	0.091	0.273	0.220	0.138
7	–	–	–	–	–	–	0.095	0.174	0.001
8	–	–	–	–	–	–	–	0.497	0.174
9	–	–	–	–	–	–	–	–	0.279
SERPINA3									
1	0.337	0.314	0.910	0.397					
2	–	0.921	0.368	0.571					
3	–	–	0.353	0.506					
4	–	–	–	0.402					

TABLE 4. *SERPINA1* SNP Groups Related to COPD at Final P < 0.01 (Omnibus Test)

No. of SNPs in group	1	2	3	4	6	7	8	10	Exact P-value (1,000 permutations)	Exact P-value (10,000 permutations)
5	1		3	4		7		10	0.004	0.0059
5	1		3	4			8	10	0.006	0.0060
6	1		3	4		7	8	10	0.003	0.0038
6	1	2	3	4			8	10	0.007	0.0044
7	1	2	3	4		7	8	10	0.006	0.0039
7	1	2	3	4	6		8	10	0.013	0.0091
8	1	2	3	4	6	7	8	10	0.009	0.0075

TABLE 5. *SERPINA1* Risk Factor Haplotypes

	SNP No												
Haplotypes	1	2	3	4	6	7	8	10	% in COPD	% in controls	P-value[a]	Odds ratio	95% CI
5-SNP	T		C	A			C	G	1.61%	0.26%	0.0002	6.2	2.5–17.9
6-SNP	T	T	C	A			C	G	1.55%	0.06%	0.0007	29.8	4.2–210.8
6-SNP	T		C	A		A	C	G	1.45%	0.25%	0.0003	6.3	2.3–17.3
7-SNP	T	T	C	A		A	C	G	1.40%	0.06%	0.0009	26.2	3.8–181.4
7-SNP	T	T	C	A	T		C	G	1.46%	0.04%	0.0024	44.5	3.9–514.4
8-SNP	T	T	C	A	T	A	C	G	1.32%	0.03%	0.0046	50.3	3.3–758.7

[a]P-values, odds ratio, and confidence intervals are corrected by logistic regression for age, gender, center, and smoking.
CI, confidence interval.

(2/5 SNPS, 2/6 SNPs, 2/7 SNPs, and 1/8 SNPs (Table 4)). There were no significant (P < 0.01) SNP groups of sizes 10, nine, four, three, or two. It was of note that none of the groups consisted of contiguous SNPs only.

The second stage of analysis for *SERPINA1* identified six haplotypes that were risk factors for COPD. These consisted of one 5-SNP, two 6-SNP, two 7-SNP, and one 8-SNP haplotypes (Table 5). It was of note that regardless of haplotype size, the SNP variant present at each SNP site was the same.

For *SERPINA3* the omnibus analysis indicated that no SNP group differed significantly (P < 0.01) between COPD cases and controls. As recommended by SAS Institute Inc. (Cary, NC), no further haplotype analysis was carried out on this gene.

Population Stratification

The overall chi-squared value for the 20 SNPs in unrelated genes of 8.584 with 20 degrees of freedom was not significant (P = 0.9872). We therefore have no evidence of stratification in our data [Pritchard and Rosenberg, 1999]. Our use of cases and controls for each recruitment center also minimized the potential for stratification.

Multiple Testing

The consistency of the *SERPINA1* SNP combinations that contribute to risk, combined with the strength of the association and the absence of any association with *SERPINA3* haplotypes, make it very unlikely that these results were a spurious outcome of multiple testing.

DISCUSSION

We identified six haplotypes in *SERPINA1* that were risk factors for COPD (Table 5). They included one 5-SNP, two 6-SNP, two 7-SNP, and one 8-SNP haplotypes, all of which have a common 5-SNP backbone. This is suggestive of a basic 5-SNP risk-associated

haplotype with specific variants at the other sites incrementally increasing the odds ratio. None of the at-risk haplotypes are common to the Z variant [Chappell et al., 2004]. While the haplotype frequencies are relatively small, the differences between cases and controls are highly significant, and for a common complex disease such as COPD they constitute a major risk.

When the same staged process was applied to the analysis of *SERPINA3* SNPs, none of the potential SNP combinations were found to be associated with disease. This provides strong evidence that variation in *SERPINA3* is not associated with significant risk of development of COPD, though a minor effect on phenotype cannot be excluded.

In assessing these results, the potential effects of population stratification and false positives must be considered. We have no evidence of stratification in our data. Our use of cohorts with controls from the same population also confers a high degree of confidence that stratification was not a contributing factor in our results. In addition, any imbalance in COPD and control numbers within centers was corrected for with the use of logistic regression.

Several factors can contribute to false positives in genetic studies, including coalescent effects of genetic and environmental background, population substructure, and multiple susceptibility alleles. It is also possible to miss real associations due to the heterogeneity that underlies common variants. In the light of previous evidence for involvement of the *SERPINA1* gene in susceptibility to COPD, this model-free, staged statistical approach, combined with the use of high-density SNP maps in a large sample with good phenotypic data, was designed to minimize the likelihood of false positives. Correction for major confounders should further unmask cryptic genetic backgrounds that contribute to susceptibility. However, replication in similar or larger independent population samples is required to verify true association.

There is some evidence that the SNPs contributing to the at-risk haplotype may reflect differences in function. We previously showed in reporter gene studies that haplotypes containing *SERPINA1* SNPs 1 and 2 are functional [Chappell et al., 2004] and that the minor allele variant for SNP 10 is associated with blunted responses to cytokines in a cell transfection system [Morgan et al., 1993, 1997]. Thus, three of the five SNPs common to all risk factor haplotypes occur in regions that have functional significance. This strongly suggests the possibility that the susceptibility haplotypes may reflect differences in *SERPINA1* expression, either in the basal state or in response to cytokines such as interleukin-6 and oncostatin-M. Cytokines can increase expression of *SERPINA1* up to 100-fold in lung epithelial cells [Boutten et al., 1998]. Such stimulated local production may be important for controlling proteolysis in the lung in response to smoking. Although we cannot completely exclude the possibility that the haplotypes of *SERPINA1* are in linkage disequilibrium with other neighboring genes, that seems unlikely given the role of this gene in COPD, and previous data on linkage disequilibrium for this gene [Chappell et al., 2004].

Many previously reported studies on genetic susceptibility in COPD were hampered by relatively small population sizes and difficulties in replication [Joos et al., 2002]. We have identified new susceptibility haplotypes of *SERPINA1* that are independent of severe deficiency, and confer a similar or higher risk of COPD than known genetic risk factors in the gene. By comparison, when corrected for confounders, no increased risk for COPD was found with the S variant, which results in mild deficiency. This is consistent with previous reports [Sandford et al., 1999].

The haplotypes of *SERPINA1* identified in this study occur at a frequency similar to that of the Z variant, but have higher odds ratios for disease than possession of the Z allele. These findings highlight the role of specific haplotypes of *SERPINA1* that confer risk to COPD, and further emphasize the key role of *SERPINA1* in this disease.

ACKNOWLEDGMENTS

We thank Ann Hann (Bristol), Breda Callaghan and Gemma Hogan (Dublin), Joyce Barr (Edinburgh), and Clara Kolster-Bijdevaate (Leiden) for assisting with recruitment; G. Catapano, E. Fornai, and C. Carli for clinical and technical assistance in Pisa; Erika Daly (Dublin) for assistance in data analysis; and Drs. B. Veldhuisen and J.J. Houwing (Leiden) for comments on data analysis.

REFERENCES

Boutten A, Venembre P, Seta N, Hamelin J, Aubier M, Durand G, Dehoux MS. 1998. Oncostatin M is a potent stimulant of alpha$_1$-antitrypsin secretion in lung epithelial cells: modulation by transforming growth factor-beta and interferon-gamma. Am J Resp Cell Mol Biol 18:511–520.

Brand PLP, Quanjer PH, Postma DS, Kerstjens HAM, Koeter GH, Dekhuijzen PNR, Sluiter HJ. 1992. Interpretation of bronchodilator response in patients with obstructive airways disease. Thorax 47:429–436.

Brantly M, Nukiwa T, Crystal RG. 1988. Molecular basis of α_1-antitrypsin deficiency. Am J Med 84:13–31.

Cavarra E, Martorana PA, Gambelli F de Santi M, van Even P, Lungarella G. 1996. Neutrophil recruitment into the lungs is associated with increased lung elastase burden, decreased lung elastin, and emphysema in alpha$_1$-proteinase inhibitor-deficient mice. Lab Invest 75:273–280.

Chappell S, Guetta-Baranes T, Batowski K, Yiannakis E, Morgan K, O'Connor C, MacNee W, Kalsheker N. 2004. Haplotypes of the alpha$_1$-antitrypsin gene in healthy controls and Z deficiency patients. Hum Mut 24:535–536.

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, Silverman EK. 2003. Genetics of chronic obstructive pulmonary disease. Semin Resp Crit Care Med 24P:157–159.

Fletcher C, Peto R. 1977. The natural history of chronic airflow obstruction. Br Med J 1:1645–1648.

Gulsvik A. 1999. Mortality and prevalence of chronic obstructive pulmonary disease in different parts of Europe. Monaldi Arch Chest Dis 54:160.

Hersh CP, Dahl M, Ly NP, Berkey CS, Nordestgaard BG, Silverman EK. 2004. Chronic obstructive pulmonary disease in alpha$_1$-antitrypsin PI MZ heterozygotes: a meta-analysis. Thorax 59: 843–849.

Joos L, Pare PD, Sandford AJ. 2002. Genetic risk factors for chronic obstructive pulmonary disease. Swiss Med Wkly 132: 27–37.

Karlinsky JB, Snider GL. 1978. Animal models of emphysema. Am Rev Respir Dis 117:1109–1133.

Mittman C, Barbella T, Lieberman J. 1973. Alpha$_1$-antitrypsin deficiency as an indicator of susceptibility to pulmonary disease. J Occup Med 15:33–38.

Morgan K, Scobie G, Marsters P, Kalsheker NA. 1993. Point mutation in a 3′ flanking sequence of the α_1-antitrypsin gene

associated with chronic respiratory disease occurs in a regulatory sequence. Hum Mol Genet 2:253–257.

Morgan K, Scobie G, Marsters P, Kalsheker NA. 1997. Mutation in an alpha$_1$-antitrypsin enhancer results in an interleukin-6 deficient acute-phase response due to loss of cooperatively between transcription factors. Biochim Biophys Acta 1362: 67–76.

Poller W, Meisen C, Olek K. 1990. DNA polymorphisms of the alpha-1-antitrypsin gene region in patients with chronic obstructive pulmonary disease. Eur J Clin Invest 20:1–7.

Poller W, Faber JP, Scholz S, Weidinger S, Bartholome K, Olek K, Eriksson S. 1992. Mis-sense mutation of α_1-anti-chymotrypsin gene associated with chronic lung disease. Lancet 339:1538.

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

Reeves EP, Lu H, Jacobs HL, Messina CGM, Bolsover S, Gabella G, Potina EO, Warley A, Roes J, Segal AW. 2002. Killing activity of neutrophils is mediated through activation of proteases by Kf flux. Nature 416:291–297.

Samilchuk E, DeSouza B, Voevodin A, Al-Awadi S. 1997. Taq 1 polymorphism in the 3′ flanking region of the PI gene among Kuwaiti Arabs and Russians. Dis Markers 13:87–92.

Sandford AJ, Spinelli JJ, Weir TD, Pare PD. 1997. Mutation in the 3′ region of the α_1-antitrypsin gene and chronic obstructive pulmonary disease. J Med Genet 34:874–875.

Sandford AJ, Chagani T, Weir TD, Pare PD. 1998. Alpha(1)-antichymotrypsin mutations in patients with chronic obstructive pulmonary disease. Dis Markers 13:257–260.

Sandford AJ, Weir TD, Spinelli JJ, Pare PD. 1999. Z and S mutations of the alpha1-antitrypsin gene and the risk of chronic obstructive pulmonary disease. Am J Respir Cell Mol Biol 20: 287–291.

Silverman EK, Chapman HA, Drazen JM, Weiss ST, Rosner B, Campbell EJ. 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.

Zaykin DV, Westfall PH, Young SS, Karnoub MA, Wagner MJ, Ehm MG. 2002. Testing associations of statistically inferred haplotypes with discrete and continuous traits in samples of unrelated individuals. Hum Hered 53:79–91.

Zhang K, Jin L. 2003. HaploBlockFinder: haplotype block analyses. Bioinformatics 10:1300–1301.

Zhao JH, Curtis D, Sham PC. 2000. Model-free analysis and permutation tests for allelic associations. Hum Hered 50: 133–139.