Eur Respir J 2008; 32: 25–34
DOI: 10.1183/09031936.00040307

Association of genetic variations in the *CSF2* and *CSF3* genes with lung function in smoking-induced COPD

J-Q. He*, K. Shumansky*, J.E. Connett#, N.R. Anthonisen¶, P.D. Paré* and A.J. Sandford*

ABSTRACT: Granulocyte-macrophage colony-stimulating factor (CSF), also known as CSF2, and granulocyte CSF, also known as CSF3, are important survival and proliferation factors for neutrophils and macrophages. The objective of the present study was to determine whether single nucleotide polymorphisms (SNPs) of *CSF2* and *CSF3* are associated with lung function in smoking-induced chronic obstructive pulmonary disease.

In total, five SNPs of *CSF2* and *CSF3* were studied in 587 non-Hispanic white subjects with the fastest (n=281) or the slowest (n=306) decline of lung function selected from among continuous smokers in the National Heart, Lung, and Blood Institute Lung Health Study (LHS). These SNPs were also studied in 1,074 non-Hispanic white subjects with the lowest (n=536) or the highest (n=538) baseline lung function at the beginning of the LHS.

An increase in the number of *CSF3* -1719T alleles was significantly associated with protection against low lung function (odds ratio 0.73, 95% confidence interval 0.56–0.95), and was still significant after adjustment for multiple comparisons. There was also a significant association of a *CSF3* haplotype with baseline levels of forced expiratory volume in one second. No association was found for *CSF2* SNPs and lung function, nor was there evidence of epistasis.

In conclusion, genetic variation in colony-stimulating factor 3 is associated with cross-sectionally measured lung function in smokers.

KEYWORDS: Chronic obstructive pulmonary disease, forced expiratory volume in one second, genetic polymorphism, granulocyte colony-stimulating factor, granulocyte-macrophage colony-stimulating factor, lung function

Chronic obstructive pulmonary disease (COPD) is a complex genetic/environmental disorder, which is characterised by airflow obstruction that is not fully reversible and by a chronic persistent inflammatory process. The degree of airflow obstruction defines disease severity, which is quantified by post-bronchodilator forced expiratory volume in one second (FEV_1) calculated as a percentage of a predicted value. Genetic factors contribute to both the level and decline of lung function. There is evidence to suggest that genetic factors account for 28.0–51.5% of the variability in cross-sectional FEV_1 [1–3] and for 18% of the variability of longitudinal change in lung function in smokers [4]. The inflammatory process is a complex interaction between many inflammatory cells. Among these cells, neutrophils and macrophages play important roles by releasing proteinases that break down connective tissue in the lung parenchyma, resulting in emphysema.

Granulocyte-macrophage colony-stimulating factor (CSF), also known as CSF2, is an important survival, proliferation and differentiation factor of the progenitor cells for neutrophils and macrophages. Granulocyte CSF, also known as CSF3, is specific for granulocytes. The *CSF2* and *CSF3* genes (located at 5q31.1 and 17q11.2–q12, respectively) were selected as candidates for studies of decline and cross-sectional level of lung function in COPD patients for the following reasons. First, CSF2 and CSF3 can induce the expression of pro-inflammatory cytokines and thereby enhance the inflammatory response. It was shown that levels of CSF2 in serum and bronchoalveolar lavage fluid (BALF), along with numbers of total cells and polymorphonuclear cells in the BALF, were increased in bronchitic

AFFILIATIONS
*The James Hogg iCAPTURE Centre for Cardiovascular and Pulmonary Research, St Paul's Hospital, University of British Columbia, Vancouver, BC, and
¶Faculty of Medicine, University of Manitoba, Winnipeg, MB, Canada.
#Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN, USA.

CORRESPONDENCE
A.J. Sandford
UBC James Hogg iCAPTURE Centre for Cardiovascular and Pulmonary Research
St Paul's Hospital
1081 Burrard Street
Vancouver
BC
Canada
V6Z 1Y6
Fax: 1 6048068351
E-mail: asandford@mrl.ubc.ca

Received:
April 05 2007
Accepted after revision:
February 28 2008

SUPPORT STATEMENT
This work was supported by grants from the Canadian Institutes of Health Research and National Institutes of Heath Grant 5R01HL064068-04. The Lung Health Study was supported by contract N01-HR-46002 from the Division of Lung Diseases of the National Heart, Lung, and Blood Institute (Bethesda, MD, USA).

STATEMENT OF INTEREST
None declared.

European Respiratory Journal
Print ISSN 0903-1936
Online ISSN 1399-3003

patients during exacerbations [5]. It was also reported that CSF3 expression in the lung correlated with severity of pulmonary neutrophilia in acute respiratory distress syndrome [6]. Secondly, it has been shown that polymorphisms and haplotypes of the *CSF2* gene are associated with the prevalence of asthma and other atopic diseases [7–9]. COPD and asthma share a common diathesis according to the ''Dutch hypothesis'' [10, 11], and atopy is a risk factor for COPD [12]. The association of a single nucleotide polymorphism (SNP) in *CSF3* with a significant increase in granulocytes among workers exposed to benzene was also reported [13]. Thirdly, a recent study directly linked the CSF2/CSF3 ratio with lung function in cystic fibrosis patients, which suggested that the interaction between CSF2 and CSF3 contributes to lung function in those patients [14].

The current authors hypothesised that *CSF2* and *CSF3* polymorphisms and their interactions would influence the decline of FEV1 and/or the cross-sectional level of FEV1 in smokers with mild-to-moderate airflow obstruction from the Lung Health Study (LHS) cohort. The LHS, sponsored by the National Heart, Lung, and Blood Institute (NHLBI; Bethesda, MD, USA), was a clinical trial of smoking intervention and bronchodilator treatment on the progression of COPD [15]. The dataset provides an excellent opportunity to explore the impact of genetic polymorphisms and their interaction on longitudinal decline and/or the cross-sectional level of FEV1 % predicted, as previously described [16–22].

METHODS

Study subjects

The LHS recruited a total of 5,887 smokers aged 35–60 yrs with spirometric evidence of mild-to-moderate lung function impairment from 10 North American medical centres. From the LHS cohort, two nested case–control studies were designed in order to study genetic determinants of rate of FEV1 decline and cross-sectional level of FEV1. Based on the rate of decline of FEV1 during 5 yrs of follow-up study, using arbitrary cut-off points of FEV1 % pred decrease of $\geqslant$3.0% per yr and increase of $\geqslant$0.4% per yr for rapid decliners and nondecliners, respectively, the 287 non-Hispanic white subjects with the highest rate of decline of lung function (fast decline group) and the 308 non-Hispanic white subjects with the slowest rate of decline of lung function (nondecline group) were selected from 3,216 continuous smokers during the first 5 yrs of follow-up. The rationale for selecting approximately the 300 highest and 300 lowest phenotypic subjects was as follows: 1) this approach has the advantage of reducing cost while keeping satisfactory statistical efficiency when compared with the full-cohort approach [23, 24]; 2) the common disease–common variant hypothesis was suggested in the late 1990s, and states that disease-susceptibility alleles of common diseases will be present at high frequencies [25–27]; and 3) this sample size has adequate power to detect common genetic risk variants, as shown previously [28]. From all remaining LHS subjects, non-Hispanic white subjects with the highest post-bronchodilator

TABLE 1 Tagging (tag) single nucleotide polymorphism (SNP) selection using the LDSelect program, and nomenclature of the SNPs

Gene	Bin#	SNP ID	SNP	Position in reference sequence¶	Position in gene+	Position in protein	Notes
CSF2	1	rs2069614	T/C	69	-1916 (promoter)		
	1	rs2069616	A/G	545	-1440 (promoter)		Genotyped
	1	rs1469149	A/C	1310	-675 (promoter)		
	1	rs743564	T/C	3347	1363 (third intron)		
	2	rs25881	C/T	3606	1622 (third intron)		Genotyped as a replacement for 1944
	2	rs25882	T/C	3928	1944 (fourth exon)	I117T	Genotype failed
	2	rs25883	G/A	4400	2416 (3′ flanking region)		
	2	rs25884	A/G	4706	2722 (3′ flanking region)		
	2	rs27438	G/A	5723	3739 (3′ flanking region)		
CSF3	1	rs2227315	A/G	28	-2012 (promoter)		
	1	rs2227322	G/C	1980	-60 (promoter)		
	1	rs1042658	T/C	4215	2176 (3′ UTR)		Genotyped
	1	rs2512146	G/T	5168	3129 (3′ flanking region)		
	2	rs2227319	G/A	1158	-882 (promoter)		Genotyped
	2	rs2227321	G/C	1607	-433 (promoter)		
	2	rs25645	G/A	3456	1417 (fifth exon)	L185L	
	2	rs2227333	C/G	4480	2441 (3′ flanking region)		
	3	rs2227316	C/T	321	-1719 (promoter)		Genotyped
	3	rs2827	C/T	4050	2011 (3′ UTR)		

All SNPs with minor allele frequency >5% are tagSNPs in this example. One SNP was genotyped in each bin. CSF: colony-stimulating factor; UTR: untranslated region. #: sites are ordered by linkage disequilibrium, with sites showing similar patterns of genotype put into the same bin; ¶: position in sequence AF373868 for *CSF2* and in sequence AF388025 for *CSF3*; +: numbered by denoting the first nucleotide of the initiator methionine codon as +1 (position 1,985 in sequence AF373868 of *CSF2* and position 2,040 in sequence AF388025 of *CSF3*).

TABLE 2 TaqMan® primer and probe sequences

SNP	Primers		Probes	
	Direction	Sequence	Allele	Sequence#
CSF2 -1440	F	AACTCCCACAGTACAGGGAAACTG	G	6FAM-ACTCAGGCCACAGTG-MGBNFQ
	R	CAGAGAGCAGGTGGAGTTCATG	A	VIC-CTCAGACCACAGTGC-MGBNFQ
CSF3 -1719	F	GCAATGAGCGAAACTCCATCTC	C	VIC-CCCACCCTCTACTC-MGBNFQ
	R	TGATGTGGCCCAGCTCTGTAC	T	6FAM-CCCACTCTCTACTCC-MGBNFQ
CSF3 -882	F	CAGCCCGTGTCCACTTCAA	G	6FAM-ACGTGACTTCCCTGGT-MGBNFQ¶
	R	TTGGAACTGCGGGATTGG	A	VIC-ACACGTGATTTCC-MGBNFQ¶
CSF3 2176	F	CAGGTGCCTGGACATTTGC	C	6FAM-CAGTCCCCGTCCAGC-MGBNFQ¶
	R	GTCTGCTCCCTCCCACATC	T	VIC-CAGTCCCCATCCAGC-MGBNFQ¶

SNP: single nucleotide polymorphism; CSF: colony-stimulating factor; F: forward; R: reverse; 6FAM and VIC: reporter fluorophores; MGBNFQ: minor groove-binding nonfluorescent quencher. #: underlining indicates SNP position; ¶: probes are designed to the reverse strand.

FEV1 % pred (high function group, n=484) and the lowest post-bronchodilator FEV1 % pred (low function group, n=468) at the beginning of the LHS were selected. Arbitrary cut-off points of FEV1 % pred $\geqslant$88.9% and $\leqslant$67.0% were used for the high and low lung function groups, respectively. Since 144 subjects from the rate of decline study groups had baseline lung function within one of the limits that defined the cross-sectional groups (58 in the high function group and 86 in the low lung function group), they were also analysed in the study of cross-sectional FEV1. Thus, there were 542 and 554 subjects in the high and low lung function groups, respectively. Informed consent was obtained from all participants and the investigation received the approval of the Providence Health Care Research Ethics Board (Vancouver, BC, Canada).

Tagging SNP selection

The *CSF2* and *CSF3* SNP discovery data were downloaded from SeattleSNPs NHLBI Program for Genomic Applications (PGA), University of Washington and Fred Hutchinson Cancer Research Center (Seattle, WA, USA) [29]. From all SNPs identified in the 23 unrelated European-American samples from the Centre d'Étude Polymorphisme Humain (CEPH) family panel, a set of tagging (tag) SNPs was chosen for each gene using the LDSelect program developed by CARLSON *et al.* [30]. A linkage disequilibrium (LD) threshold of $r^2>0.64$ and minor allele frequency of 5% were used. Initially, two SNPs located at -1440A/G and 1944T/C (I117T) in the *CSF2* gene were selected; however, testing for the 1944T/C SNP could not be established by the TaqMan® assay (Applied Biosystems, Foster City, CA, USA), and a restriction fragment length polymorphism (RFLP) PCR assay for the same SNP showed that PCR amplification failed for some samples. Therefore, 1944T/C was replaced with an alternative SNP, 1622C/T. In the *CSF3* gene, three SNPs located at -1719C/T, -882G/A and 2176T/C were selected and genotyped. TagSNP selection and the nomenclature of the SNPs are presented in table 1.

Genotyping

All SNPs except *CSF2* 1622C/T were genotyped in 384-well plates with a total volume of 5 μL by the TaqMan® 5′ exonuclease assay using primers and probes supplied by Applied Biosystems on an ABI Prism 7900HT Sequence Detection System (Applied Biosystems). Probe and primer sequences for each assay are listed in table 2. Major and minor

TABLE 3 Distribution of demographic characteristics for the longitudinal change study of forced expiratory volume in one second (FEV1)

	Fast decliners	Nondecliners	p-value
Subjects (M/F) n	281 (164/117)	306 (204/102)	0.038
Age yrs	49.51±0.38	47.61±0.39	0.0006
Smoking history pack-yrs	42.86±1.14	38.38±1.04	0.004
ΔFEV1·yr^{-1} % pred#	-4.14±0.06	1.08±0.04	<0.0001
Baseline FEV1 % pred#	72.6±0.53	75.7±0.46	<0.0001

Data are presented as mean±SE, unless otherwise stated. M: male; F: female; ΔFEV1·yr^{-1}: change in FEV1 per year over a 5-yr period; % pred: % predicted. #: pre-bronchodilator.

TABLE 4 Distribution of demographic characteristics for the cross-sectional study of forced expiratory volume in one second (FEV1)

	High FEV1	Low FEV1	p-value
Subjects (M/F) n	536 (354/182)	538 (332/206)	0.139
Age yrs	46.24±0.30	50.69±0.26	<0.0001
Smoking history pack-yrs	35.32±0.77	45.16±0.81	<0.0001
ΔFEV1·yr^{-1} % pred			
Pre-bronchodilator	-0.55±0.07	-1.27±0.08	<0.0001
Post-bronchodilator	-0.75±0.06	-0.79±0.08	<0.722
Baseline FEV1 % pred			
Pre-bronchodilator	86.48±0.13	61.08±0.18	<0.0001
Post-bronchodilator	91.80±0.10	62.61±0.14	<0.0001

Data are presented as mean±SE, unless otherwise stated. M: male; F: female; ΔFEV1·yr^{-1}: change in FEV1 per year over a 5-yr period; % pred: % predicted.

TABLE 5 The relationship between two linkage disequilibrium-selected tagging (tag) single nucleotide polymorphisms (SNPs) and haplotypes resolved by them

Haplotype	Frequency %	SNP position in the *CSF2* gene								
		-1916	-1440#	-675	1363	1622#	1944	2416	2722	3739
1	41.30	W	W	W	W	W	W	W	W	W
2	34.78	M	M	M	M	W	W	W	W	W
3	17.39	W	W	W	W	M	M	M	M	M
4	4.35	W	W	W	M	M	M	W	M	W
5	2.17	M	W	M	M	M	W	W	W	W

CSF: colony-stimulating factor; W: major alleles; M: minor alleles. #: genotyped tagSNPs.

probes were labelled with 5′ FAM or 5′ VIC fluorophores as reporters (Applied Biosystems). For each SNP genotyping, ⩽47 DNA samples of the CEPH panel with sequencing information available from the SeattleSNPs PGA were included as quality controls. All genotype results from the TaqMan® assay were consistent with sequencing results for all CEPH DNA samples that have sequencing information available in the SeattleSNPs database. No discrepancies were detected in the 10% of the randomly selected samples that were genotyped in duplicate.

The *CSF2* 1622C/T polymorphism was detected by an RFLP-PCR method using the following primers flanking the polymorphic region: 5′-AAGGAAGGGAGGCTACTTGG-3′ (sense) and 5′-GTTCCCCAAGGAGTGCATAG-3′ (antisense). Amplification products were digested by the *Blp*I restriction enzyme. *Blp*I produced 116-bp and 133-bp fragments when 1622T was present, but did not digest the 249bp PCR product when *CSF2* 1622C was present. The genotyping method was confirmed by sequencing 10 samples with three different genotypes. Sequencing was performed on an ABI 3100 16-capillary automated genetic analyser (Applied Biosystems) using the same primers as in the PCR reaction that produced the sequence template.

Statistical analysis

Hardy–Weinberg equilibrium tests and LD estimation were performed using the genetics package for R [31]. All single-locus association tests were performed in R. The codominant

TABLE 6 Individual single nucleotide polymorphism (SNP) associations in the studies of longitudinal change of forced expiratory volume in one second (FEV_1) and cross-sectional level of FEV_1

SNP	Genotype	Longitudinal change study#		Cross-sectional study					
		Rapid decliners	Nondecliners	Low function	High function	Codominant model		Additive model	
						Unadjusted p-value	Adjusted p-value¶	Unadjusted p-value	Adjusted p-value¶
***CSF2* -1440**	AA+	100 (35.7)	101 (33.0)	166 (32.2)	180 (34.2)				
	AG	124 (44.3)	152 (49.7)	252 (48.9)	259 (49.2)	NS	NS	NS	NS
	GG	56 (20.0)	53 (17.3)	97 (18.8)	88 (17.7)				
***CSF2* 1622**	CC+	194 (69.3)	210 (68.6)	353 (68.2)	340 (64.4)				
	CT	74 (26.4)	83 (27.1)	140 (27.0)	171 (32.4)	0.092	NS	NS	NS
	TT	12 (4.3)	13 (4.2)	25 (4.8)	17 (3.2)				
***CSF3* -1719**	CC+	199 (71.1)	214 (70.2)	388 (74.3)	364 (69.2)				
	CT	71 (25.4)	78 (25.6)	124 (23.8)	147 (28.0)	NS	0.059	0.054	0.018
	TT	10 (3.6)	13 (4.3)	10 (1.9)	15 (2.9)				
***CSF3* -882**	GG+	117 (41.8)	112 (36.7)	202 (39.1)	224 (43.2)				
	GA	124 (44.3)	155 (50.8)	234 (45.3)	233 (45.0)	NS	NS	0.059	0.092
	AA	39 (13.9)	38 (12.5)	81 (15.7)	61 (11.8)				
***CSF3* 2176**	CC+	97 (35.4)	105 (34.8)	192 (36.8)	185 (35.3)				
	CT	130 (47.4)	160 (53.0)	247 (47.3)	259 (49.3)	NS	NS	NS	NS
	TT	47 (17.2)	37 (12.3)	83 (15.9)	80 (15.3)				

Data are presented as n (%). CSF: colony-stimulating factor; NS: nonsignificant. #: unadjusted p-values and p-values adjusted for age, sex, smoking pack-yrs and research centre for the codominant and additive models were all nonsignificant; ¶: adjusted for age, sex, smoking pack-yrs, research centre and the rate of decline of FEV_1 % predicted post-bronchodilator; +: homozygote for the major allele.

TABLE 7 Individual single nucleotide polymorphism (SNP) associations with forced vital capacity (FVC) in the study of longitudinal change of forced expiratory volume in one second

SNP	Subjects n	Genotype	FVC % pred	F-value#	p-value#
***CSF2* -1440**	413	CC	97.92±0.53		
	149	CT	97.49±0.89	0.675	0.513
	23	TT	100.65±2.57		
***CSF2* 1622**	229	GG	97.39±0.72		
	279	GA	97.99±0.65	0.906	0.406
	77	AA	99.27±1.20		
***CSF3* -1719**	201	AA	98.51±0.75		
	276	AG	97.48±0.67	0.528	0.590
	109	GG	98.03±1.00		
***CSF3* -882**	404	CC	97.44±0.52		
	157	CT	99.03±0.93	1.220	0.302
	25	TT	98.99±2.42		
***CSF3* 2176**	202	CC	97.32±0.79		
	290	CT	98.94±0.60	3.466	0.033
	84	TT	95.65±1.21		

Data are presented as mean± SE, unless otherwise stated. CSF: colony-stimulating factor; % pred: % predicted. #: overall comparisons of the differences in FVC % pred among the three genotype groups.

and additive models were tested first and, if there was a significant association, the dominant and recessive models were tested additionally, to see if those models fitted better. If the cell counts were low, significance was assessed by permutation tests. In a codominant model, a heterozygote shows the phenotypic effects of both alleles fully and equally. The three genotypic categories of each SNP in the case and control groups constitute a 2×3 contingency table and the analysis does not provide any sense of ordering across the three genotypes. This type of analysis is also called a general genetic model [32]. In a dominant model, one copy of the minor allele increases disease risk. The homozygotes and heterozygotes for the minor allele are compared as a group with homozygotes for the major allele [32]. In a recessive model, two copies of the minor allele are required to increase disease risk. The homozygotes for the minor allele are compared with heterozygotes and homozygotes for the major allele as a group. In an additive model, there is r-fold increased disease risk for heterozygotes compared with the homozygotes for the major allele, and 2r-fold increased disease risk for the homozygotes for the minor allele compared with the homozygotes for the major allele [32]. The Armitage trend test [33] was used to test an additive effect of the allele. In both the FEV1-decline study and the cross-sectional FEV1 study, in addition to crude analysis by Chi-squared tests using 2×3 contingency tables, multivariate logistic regression analyses were also used to control for potential confounders that might influence the rate of decline of lung function or the cross-sectional FEV1 level. In the FEV1-decline study, multivariate logistic regression was used to adjust for confounding factors such as age, sex, smoking pack-yrs and research centre. In the cross-sectional FEV1 study, multivariate logistic regression was used to adjust for the same confounding factors plus the rate of decline of FEV1. Although other phenotypes, such as forced vital capacity (FVC) % pred and FEV1/FVC ratio, were not the primary phenotypes due to study design, associations of those phenotypes with single SNPs were also analysed, using one-way ANOVA if the data were normally distributed or a Spearman's rank test if the data were not normally distributed in the study groups.

The effective number (n_e) of haplotypes from SNPs with minor allele frequency $\geqslant$5% was calculated by the following equation, where p_i is the frequency of haplotype i [30].

$$n_e = \frac{1}{\sum_i p_i^2} \quad (1)$$

The n_e of haplotypes weights the number of haplotypes by frequency, with common haplotypes more heavily weighted.

Correction for multiple tests of SNPs in LD in each gene was carried out on the basis of the spectral decomposition (SpD) of matrices of pairwise LD between SNPs using SNPSpD [34, 35]. This method provides a useful alternative to the very conservative Bonferroni correction. Haplotype association was tested using the hapassoc package for R [31]. This software performs likelihood inference of trait associations with haplotypes and other covariates for generalised linear models, including logistic regression, and does not assume that haplotype phase is known [36]. An additive effect of haplotype on the log-odds of disease was assumed. To calculate haplotype frequencies, an Expectation Maximization algorithm from the haplo.stats package for R was used [31].

Focused interaction testing framework (FITF), was used to identify gene–gene interactions [37].

Power analysis

The power of the two studies was estimated using the two independent proportions and many proportions functions in Power Analysis and Sample Size (PASS) 2005 [38]. Plots were created in R [31] using the output from PASS.

RESULTS

Characteristics of the study groups

The characteristics of study participants are shown in tables 3 and 4. Since there was no DNA available for eight subjects in the rate of decline of FEV1 study or for 22 subjects in the cross-sectional level of FEV1 study, the numbers of participants in the two studies were 587 and 1,074, respectively.

Among nondecliners from the rate of decline of FEV1 study and among the high lung function group of the cross-sectional FEV1 study, the allele frequencies of all five SNPs did not significantly deviate from Hardy–Weinberg equilibrium (results not shown).

Haplotypes resolved with the genotyped tagSNPs

Haplotypes from SNPs with minor allele frequency $\geqslant$5% in 23 CEPH samples were inferred by use of PHylogenetics And Sequence Evolution (PHASE) 2.0 [39, 40]. The LD-selected *CSF2* tagSNPs could resolve 60% of the actual number of haplotypes (three out of five; table 5) and 87.1% (2.7 out of 3.1)

TABLE 8 Haplotype association in the cross-sectional study of forced expiratory volume in one second (FEV_1)

Gene	Haplotype[#]	Low function %	High function %	Global test p-values[¶]		p-values[+]	
				Unadjusted	Adjusted[§]	Unadjusted	Adjusted[§]
CSF2	-1440A/1622C	33.0	36.9				
	-1440G/1622C	49.7	44.3	NS	NS	NS	NS
	-1440G/1622T	17.3	20.0			NS	NS
CSF3	-1719C/-882G/2176T	39.9	39.9				
	-1719C/-882A/2176C	38.7	34.2	0.004	0.027	0.007	0.089
	-1719T/-882G/2176C	14.0	16.9			NS	0.047
	-1719C/-882G/2176C	7.5	8.9			NS	NS
	-1719C/-882G	48.4	48.8				
	-1719C/-882A	38.7	34.2	0.069	0.0427	NS	NS
	-1719T/-882G	14.0	16.9			NS	0.058
	-882G/2176T	39.9	39.9				
	-882A/2176C	38.7	34.2	0.051	0.053	NS	NS
	-882G/2176C	21.5	25.8			NS	0.081

CSF: colony-stimulating factor; NS: nonsignificant. [#]: calculated using an Expectation Maximization algorithm from the haplo.stats package for R; [¶]: comparing overall haplotype distribution between cases and controls; [+]: comparing each haplotype between cases and controls, using the most common haplotype as a reference; [§]: adjusted for age, sex, smoking pack-yrs, research centre and the rate of decline of FEV_1 % predicted post-bronchodilator.

of the n_e of haplotypes from SNPs with minor allele frequency $\geqslant$5%. For *CSF3*, 35.7% of actual haplotypes and 48.1% of effective haplotypes from SNPs with a minor allele frequency $\geqslant$5% were resolved by the three selected tagSNPs.

Individual SNP association analysis

In the FEV_1-decline study, none of the five SNPs were associated with decline of FEV_1 in codominant or additive models either before or after adjustment for confounding factors (table 6).

In the study of the cross-sectional level of FEV_1, there was a borderline association of *CSF3* -1719T with high FEV_1 levels in an additive model (p=0.054) before adjustment for confounding factors; after adjustment for confounding factors, the association was more significant (p=0.018; table 6). The odds ratios (ORs) of having one -1719T allele compared with no -1719T allele and of having two -1719T alleles compared with one -1719T allele were both 0.73, 95% confidence interval 0.56–0.95. The association of *CSF3* -1719 with FEV_1 level was adjusted for multiple testing on the basis of the SNPSpD approach [34, 35]. The significance threshold required to keep the type I error rate at 5% for *CSF3* in the present study was 0.019, based on the LD of the three SNPs studied. Therefore, the association of *CSF3* -1719 with FEV_1 level remained significant after correction for multiple comparisons.

In addition, two SNPs showed borderline associations with FEV_1 levels before adjustment for confounding factors: *CSF2* 1622 in a codominant model (comparison of the distribution of the three genotypic groups, CC, CT and TT, in the case and control groups gave p=0.092); and *CSF3* -882 in an additive model (the OR of one A allele compared with no A allele was equal to the OR of two A alleles compared with one A allele; p=0.059). However, after adjustment for confounding factors the p-values were less significant for both SNPs (table 6).

Although FVC % pred and the FEV_1/FVC ratio at the beginning of the LHS were not the primary phenotypes due to the case–control study design, exploratory analyses of single SNP associations were performed with those phenotypes. In the rate of decline group, the FVC % pred phenotype was normally distributed; therefore, a one-way ANOVA was used to investigate whether FVC % pred and the FEV_1/FVC ratio were the same among the three genotypic groups. A significant association of FVC % pred with *CSF3* 2176 was found (p=0.033; table 7): those individuals with the 2176TT genotype had a lower FVC % pred. No other significant associations were found (data not shown).

Haplotype association analysis

Haplotypes from *CSF2* or *CSF3* were not associated with decline of FEV_1 in the analysis either with or without adjustment for confounding factors (data not shown).

The results of haplotype association in the cross-sectional level of FEV_1 study are shown in table 8. The haplotypes from *CSF2* were not associated with decline of FEV_1 in the analysis either with or without adjustment for confounding factors. The three-locus *CSF3* haplotypes were associated with levels of FEV_1 in a Wald global test before adjustment for confounding factors (an overall test of haplotype distribution between cases and controls gave p=0.004), although after adjustment for confounding factors, the association became less significant (p=0.027). The frequency of the haplotype -1719T/-882G/2176C was marginally higher in the high than the low function FEV_1 group (16.9 *versus* 14.0%) when compared with the haplotype -1719C/-882G/2176T as a reference (adjusted p-value 0.047). Analysis of two-locus haplotypes (table 8) demonstrated that this marginal association was probably driven by both the -1719T allele and the 2176C allele. The frequency of the haplotype -1719C/-882A/2176C was lower in the high than the low function FEV_1 group (34.2 *versus* 38.7%)

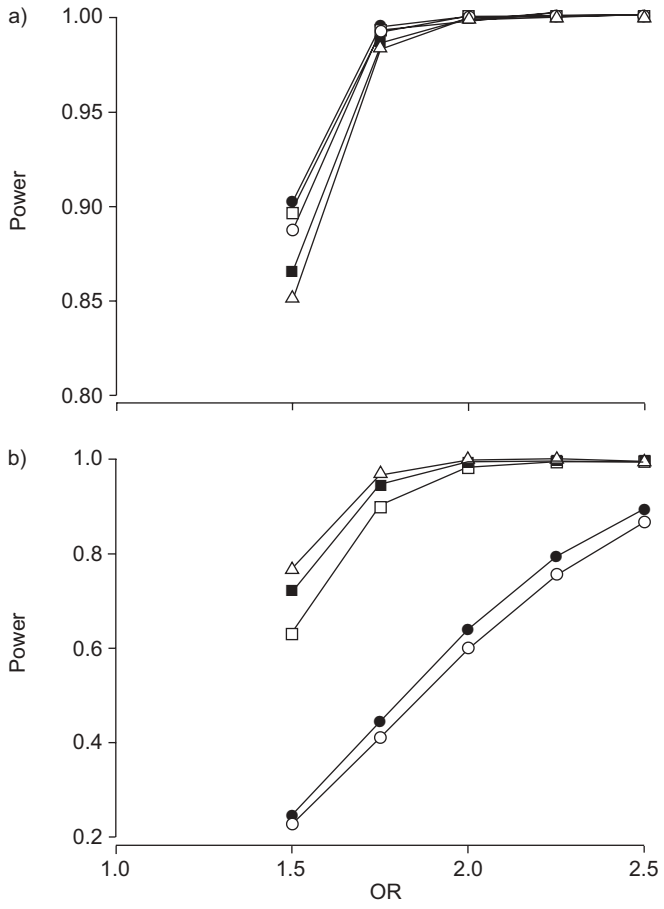

FIGURE 1. Power curves for a) dominant models and b) recessive models, for the five single nucleotide polymorphisms in the cross-sectional study of baseline forced expiratory volume in one second. Two-sided pooled Z-tests were performed, *i.e.* the proportions were pooled to compute the SE, using a significance level (probability of a type I error) of 0.05. The sample sizes for the case (low lung function) and control (high lung function) groups were 538 and 536, respectively. ○: colony-stimulating factor (*CSF*) *3* -1719; ●: *CSF2* 1662; □: *CSF3* -882; ■: *CSF3* 2176; △: *CSF2* -1440. P2 values, representing the proportion of the risk genotype group, were as follows: a) *CSF3* -1719=0.3090, *CSF2* 1662=0.3560, *CSF3* -882=0.5680, *CSF3* 2176=0.6460 and *CSF2* -1440=0.6690; b) *CSF3* -1719=0.0290, *CSF2* 1662=0.0320, *CSF3* -882=0.1180, *CSF3* 2176=0.1530 and *CSF2* -1440=0.1770. OR: odds ratio.

when compared with the haplotype -1719C/-882G/2176T as a reference, but the significance became borderline when adjusting for confounding factors (unadjusted p-value 0.007, adjusted p-value 0.089).

Gene–gene interactions

Using the FITF method, interactions of *CSF2* and *CSF3* were explored for all possible two- to four-locus models. There was no evidence of epistasis (gene–gene interaction; detailed results not shown).

Power analysis

The power of the present study was first calculated for a codominant mode of inheritance. A Chi-squared test with two degrees of freedom was used to calculate the associated power. Effect size, a measure of the magnitude of the Chi-squared value that is to be detected and a parameter needed for the power calculations, was calculated using the PASS program for each SNP and was used in the calculations. It was found that there was >80% power to detect an OR of 1.75 for both FEV1 decline and cross-sectional level studies. The power of the dominant and recessive models was tested with a 2×2 table; the proportions in the control group were set to be close to those observed with the five SNPs in the ''low'' outcome groups (*i.e.* nondecline of FEV1 group and high lung function group). Figure 1 shows the curves of power *versus* OR value for the five studied SNPs in the cross-sectional study of baseline level of FEV1 for dominant and recessive models. For the FEV1-decline study, the power was slightly less than that of the baseline FEV1 study, due to smaller sample size (data not shown).

DISCUSSION

CSF3 is a logical candidate gene for the present study due to its biological function. In a rat model, neutrophil stimulation by CSF3 aggravates ventilator-induced lung injury manifested by increased lung neutrophils and interleukin-6 expression, increased alveolar oedema on histology and reduced lung compliance [41]. In patients with acute respiratory distress syndrome, CSF3 expression level in the lung correlated with severity of pulmonary neutrophilia [6]. Recently, it was shown that the *CSF3* 2176 SNP (named exon 4-165C>T) was associated with peripheral blood granulocyte count among workers exposed to benzene [13]. Subjects with homozygous TT genotypes had significantly increased blood granulocytes compared with homozygous CC subjects (p=0.00002) [13].

The functional significance of the *CSF3* SNPs is unknown. Although no association of *CSF3* 2176 with the primary phenotypes of baseline and decline of FEV1 was found, a different SNP, *CSF3* 1719, was found to be associated with the baseline level of FEV1. Interestingly, in an exploratory analysis of individual SNPs with other phenotypes, such as FVC % pred and the FEV1/FVC ratio, a significant association of *CSF3* 2176 with FVC % pred was found (without correction for multiple comparisons). The association of the *CSF3* 2176TT genotype with lower FVC % pred is consistent with the previous report that the TT genotype was associated with higher blood granulocytes [13], since neutrophils in the lung and in the blood are important effector cells in COPD [42].

There are several explanations for these differences from previous studies, including genetic heterogeneity between different populations and differences in phenotypes studied and in tagSNP choice. It was reported previously that tagSNPs selected using the criteria of $r^2=0.64$ and minor allele frequency of 5% could resolve 76% of actual and 85% of effective haplotypes in an analysis of 100 genes [30]. However, using the same criteria, the *CSF3* tagSNPs only resolved 35.7% of the actual haplotypes and 48.1% of the effective haplotypes. If *CSF3* 2176 is not the causal SNP and there are different LD patterns in the present study population compared with that of the workers exposed to benzene [13], the functional SNP might have been missed in the current study. The fact that the present results showed that *CSF3* 1719 was associated with the baseline level of FEV1 and *CSF3* 2176 was associated with FVC % pred suggests that neither SNP is causal but may be in LD with a causal SNP that is yet to be identified.

The observation that SNPs from *CSF3* but not *CSF2* were associated with lung function has several possible explanations. First, animal studies have documented that CSF3 plays a more important role than CSF2 in regulation of neutrophil homeostasis. Dogs depleted of CSF3 by a neutralising antibody developed profound and selective neutropenia [43], whereas mice depleted of CSF2 did not show impairment of haematopoiesis [44]. In addition, *CSF3* but not *CSF2* knockout mice display chronic neutropenia [45, 46]. Secondly, in patients with acute respiratory distress syndrome, CSF3 but not CSF2 expression in the lung correlated with severity of pulmonary neutrophilia [6], which demonstrated that CSF3 also plays a more important role than CSF2 in regulation of neutrophils in human subjects. Thirdly, it was reported that dexamethasone inhibits human airway smooth muscle cell release of CSF2 but not CSF3 [47], suggesting that CSF3 and CSF2 are released through different mechanisms and, thus, may play different roles in the development of COPD.

It has been suggested that, apart from mobilising granulocytes from the bone marrow, CSF2 and CSF3 are decisive in influencing the subsequent T-helper cell (Th) type 1 or Th2 dominance of the immune response by selecting subsets of dendritic cells [14]. A recent study demonstrated that a high CSF2/CSF3 ratio was correlated with good lung function in cystic fibrosis patients with chronic *Pseudomonas aeruginosa* lung infection [14], which prompted the current analysis of gene–gene interaction. However, no significant interaction of CSF2 and CSF3 was found in the present study. There are several potential reasons for this. First, the current study might not have had enough power to detect gene–gene interaction due to the sample size and minor allele frequencies used. Secondly, cystic fibrosis with chronic *P. aeruginosa* lung infection is a Th2-dominated response [48] while COPD is a Th1-dominated response [49]. Therefore, the determinants of lung function in cystic fibrosis patients with chronic *P. aeruginosa* lung infection and in smoking-induced COPD patients are likely to be different.

There are several limitations to the present study. First, population stratification could have led to false-positive results. However, it has been reported that in the non-Hispanic white population, significant false-positive associations are unlikely to arise from population stratification, especially in well-designed, moderately-sized, case–control studies [50, 51]. Secondly, false-positive results might have arisen from multiple comparisons. Although the results of association of *CSF3* -1719 with lung function were corrected for multiple comparisons, only multiple SNPs in a single gene were taken into account. No correction for multiple genes and phenotypes was performed. Thirdly, no replication was performed by analysing a second cohort. Fourthly, no available function data support the associations that were found. Finally, the method of nested case–control study (*i.e.* using individuals from each extreme of the distribution of the phenotype of interest) has the advantages of cost reduction combined with satisfactory statistical efficiency when compared with the full cohort approach [23, 24]. However, this study design prevented analysis of baseline and decline in FEV_1 as continuous variables. Therefore, the results from the present study should be regarded only as hypothesis generating, and it will be necessary to replicate them in different studies, especially in those with a cohort design.

In conclusion, an association of the colony-stimulating factor 3 -1719C/T single nucleotide polymorphism with the baseline level of forced expiratory volume in one second was found. However, this association needs to be replicated in further studies. Moreover, additional functional study of this single nucleotide polymorphism, or other single nucleotide polymorphisms in linkage disequilibrium with it, is warranted.

REFERENCES

1 Astemborski JA, Beaty TH, Cohen BH. Variance components analysis of forced expiration in families. *Am J Med Genet* 1985; 21: 741–753.
2 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.
3 Wilk JB, Djousse L, Arnett DK, *et al.* Evidence for major genes influencing pulmonary function in the NHLBI family heart study. *Genet Epidemiol* 2000; 19: 81–94.
4 Gottlieb DJ, Wilk JB, Harmon M, *et al.* Heritability of longitudinal change in lung function. The Framingham study. *Am J Respir Crit Care Med* 2001; 164: 1655–1659.
5 Balbi B, Bason C, Balleari E, *et al.* Increased bronchoalveolar granulocytes and granulocyte/macrophage colony-stimulating factor during exacerbations of chronic bronchitis. *Eur Respir J* 1997; 10: 846–850.
6 Aggarwal A, Baker CS, Evans TW, Haslam PL. G-CSF and IL-8 but not GM-CSF correlate with severity of pulmonary neutrophilia in acute respiratory distress syndrome. *Eur Respir J* 2000; 15: 895–901.
7 Rohrbach M, Frey U, Kraemer R, Liechti-Gallati S. A variant in the gene for GM-CSF, I117T, is associated with atopic asthma in a Swiss population of asthmatic children. *J Allergy Clin Immunol* 1999; 104: 247–248.
8 Rafatpanah H, Bennett E, Pravica V, *et al.* Association between novel GM-CSF gene polymorphisms and the frequency and severity of atopic dermatitis. *J Allergy Clin Immunol* 2003; 112: 593–598.
9 He JQ, Ruan J, Chan-Yeung M, *et al.* Polymorphisms of the GM-CSF genes and the development of atopic diseases in at-risk children. *Chest* 2003; 123: Suppl. 3, 438S.
10 Sluiter HJ, Koëter GH, de Monchy JG, Postma DS, de Vries K, Orie NG. The Dutch hypothesis (chronic non-specific lung disease) revisited. *Eur Respir J* 1991; 4: 479–489.
11 Vestbo J, Prescott E. Update on the ''Dutch hypothesis'' for chronic respiratory disease. *Thorax* 1998; 53: Suppl. 2, S15–S19.
12 Weiss ST. Atopy as a risk factor for chronic obstructive pulmonary disease: epidemiological evidence. *Am J Respir Crit Care Med* 2000; 162: S134–S136.
13 Lan Q, Zhang L, Shen M, *et al.* Polymorphisms in cytokine and cellular adhesion molecule genes and susceptibility to hematotoxicity among workers exposed to benzene. *Cancer Res* 2005; 65: 9574–9581.
14 Moser C, Jensen PØ, Pressler T, *et al.* Serum concentrations of GM-CSF and G-CSF correlate with the Th1/Th2 cytokine response in cystic fibrosis patients with chronic *Pseudomonas aeruginosa* lung infection. *APMIS* 2005; 113: 400–409.

15 Anthonisen NR, Connett JE, Kiley JP, *et al.* Effects of smoking intervention and the use of an inhaled anticholinergic bronchodilator on the rate of decline of FEV1. The Lung Health Study. *JAMA* 1994; 272: 1497–1505.

16 He JQ, Connett JE, Anthonisen NR, Paré PD, Sandford AJ. Glutathione S-transferase variants and their interaction with smoking on lung function. *Am J Respir Crit Care Med* 2004; 170: 388–394.

17 He JQ, Connett JE, Anthonisen NR, Sandford AJ. Polymorphisms in the IL13, IL13RA1, and IL4RA genes and rate of decline in lung function in smokers. *Am J Respir Cell Mol Biol* 2003; 28: 379–385.

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

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

20 Joos L, He JQ, Shepherdson MB, *et al.* The role of matrix metalloproteinase polymorphisms in the rate of decline in lung function. *Hum Mol Genet* 2002; 11: 569–576.

21 He JQ, Burkett K, Connett JE, Anthonisen NR, Paré PD, Sandford AJ. Interferon-γ polymorphisms and their interaction with smoking are associated with lung function. *Hum Genet* 2006; 119: 365–375.

22 He JQ, Shumansky K, Zhang X, Connett JE, Anthonisen NR, Sandford AJ. Polymorphisms of interleukin-10 and its receptor and lung function in COPD. *Eur Respir J* 2007; 29: 1120–1126.

23 Allison DB, Heo M, Schork NJ, Wong SL, Elston RC. Extreme selection strategies in gene mapping studies of oligogenic quantitative traits do not always increase power. *Hum Hered* 1998; 48: 97–107.

24 Ernster VL. Nested case–control studies. *Prev Med* 1994; 23: 587–590.

25 Lander ES. The new genomics: global views of biology. *Science* 1996; 274: 536–539.

26 Cargill M, Altshuler D, Ireland J, *et al.* Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat Genet* 1999; 22: 231–238.

27 Chakravarti A. Population genetics–making sense out of sequence. *Nat Genet* 1999; 21: Suppl. 1, 56–60.

28 Tanaka G, Sandford AJ, Burkett K, *et al.* Tumour necrosis factor and lymphotoxin A polymorphisms and lung function in smokers. *Eur Respir J* 2007; 29: 34–41.

29 SeattleSNPS Variation Discovery Resource. http://pga.gs.washington.edu Date last accessed: October 2003. Date last updated: February 14, 2008.

30 Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am J Hum Genet* 2004; 74: 106–120.

31 The R Project for Statistical Computing. www.r-project.org Date last accessed: February 2006. Date last updated: February 8, 2008.

32 Lewis CM. Genetic association studies: design, analysis and interpretation. *Brief Bioinform* 2002; 3: 146–153.

33 Armitage P. Tests for linear trends in proportions and frequencies. *Biometrics* 1955; 11: 375–386.

34 Single Nucleotide Polymorphism Spectral Decomposition (SNPSpD). http://genepi.qimr.edu.au/general/daleN/SNPSpD/ Date last accessed: February 2006. Date last updated: October 12, 2007.

35 Nyholt DR. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am J Hum Genet* 2004; 74: 765–769.

36 Burkett K, McNeney B, Graham J. A note on inference of trait associations with SNP haplotypes and other attributes in generalized linear models. *Hum Hered* 2004; 57: 200–206.

37 Millstein J, Conti DV, Gilliland FD, Gauderman WJ. A testing framework for identifying susceptibility genes in the presence of epistasis. *Am J Hum Genet* 2006; 78: 15–27.

38 Hintze J. NCSS and PASS. Number Cruncher Statistical Systems. Kaysville, 2004. www.ncss.com Date last accessed: February 2006. Date last updated: January 14, 2008.

39 Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 2001; 68: 978–989.

40 Stephens M, Donnelly P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet* 2003; 73: 1162–1169.

41 Karzai W, Cui X, Heinicke N, *et al.* Neutrophil stimulation with granulocyte colony-stimulating factor worsens ventilator-induced lung injury and mortality in rats. *Anesthesiology* 2005; 103: 996–1005.

42 Oudijk EJ, Lammers JW, Koenderman L. Systemic inflammation in chronic obstructive pulmonary disease. *Eur Respir J* 2003; 22: Suppl. 46, 5s–13s.

43 Hammond WP, Csiba E, Canin A, *et al.* Chronic neutropenia. A new canine model induced by human granulocyte colony-stimulating factor. *J Clin Invest* 1991; 87: 704–710.

44 Dranoff G, Crawford AD, Sadelain M, *et al.* Involvement of granulocyte-macrophage colony-stimulating factor in pulmonary homeostasis. *Science* 1994; 264: 713–716.

45 Stanley E, Lieschke GJ, Grail D, *et al.* Granulocyte/macrophage colony-stimulating factor-deficient mice show no major perturbation of hematopoiesis but develop a characteristic pulmonary pathology. *Proc Natl Acad Sci USA* 1994; 91: 5592–5596.

46 Lieschke GJ, Grail D, Hodgson G, *et al.* Mice lacking granulocyte colony-stimulating factor have chronic neutropenia, granulocyte and macrophage progenitor cell deficiency, and impaired neutrophil mobilization. *Blood* 1994; 84: 1737–1746.

47 Patel HJ, Belvisi MG, Bishop-Bailey D, Yacoub MH, Mitchell JA. Activation of peroxisome proliferator-activated receptors in human airway smooth muscle cells has a superior anti-inflammatory profile to corticosteroids: relevance for chronic obstructive pulmonary disease therapy. *J Immunol* 2003; 170: 2663–2669.

48 Moser C, Kjaergaard S, Pressler T, Kharazmi A, Koch C, Høiby N. The immune response to chronic *Pseudomonas aeruginosa* lung infection in cystic fibrosis patients is predominantly of the Th2 type. *APMIS* 2000; 108: 329–335.

49 Majori M, Corradi M, Caminati A, Cacciani G, Bertacco S, Pesci A. Predominant TH1 cytokine pattern in peripheral blood from subjects with chronic obstructive pulmonary disease. *J Allergy Clin Immunol* 1999; 103: 458–462.

50 Ardlie KG, Lunetta KL, Seielstad M. Testing for population subdivision and association in four case–control studies. *Am J Hum Genet* 2002; 71: 304–311.

51 Wacholder S, Rothman N, Caporaso N. Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. *J Natl Cancer Inst* 2000; 92: 1151–1158.