ERJ Express. Published on March 19, 2008 as doi: 10.1183/09031936.00040307

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

Jian-Qing He[1], Karey Shumansky[1], John E Connett[2], Nicholas R Anthonisen[3], Peter D Paré[1], Andrew J Sandford[1]

1. The James Hogg iCAPTURE Centre for Cardiovascular and Pulmonary Research, St. Paul's Hospital, University of British Columbia, Vancouver, B.C., Canada.

2. Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, Minnesota, USA

3. Faculty of Medicine, University of Manitoba, Winnipeg, Manitoba, Canada

ADDRESS FOR CORRESPONDENCE:

Dr. A. J. Sandford

UBC James Hogg iCAPTURE Centre for Cardiovascular and Pulmonary Research, St. Paul's Hospital, 1081 Burrard Street, Vancouver, B.C., Canada. V6Z 1Y6.

TEL: (604) 806-9008

FAX: (604) 806-8351

Email: asandford@mrl.ubc.ca

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.

Running title: CSF2 and CSF3 polymorphisms, lung function

Word count: Body text 3,858; Abstract 200

ABSTRACT

Background: Granulocyte-macrophage colony-stimulating factor (CSF2) and G-CSF (CSF3) are important survival and proliferation factors for neutrophils and macrophages. The objective of this study was to determine whether single nucleotide polymorphisms (SNPs) of *CSF2* and *CSF3* are associated with lung function in smoking induced COPD.

Methods: Five SNPs of *CSF2* and *CSF3* were studied in 587 non-Hispanic whites with the fastest (n=281) or the slowest (n=306) decline of lung function selected from among continuous smokers in the NHLBI Lung Health Study (LHS). These SNPs were also studied in 1074 non-Hispanic whites with the lowest (n=536) or the highest (n=538) baseline lung function at the beginning of the LHS.

Results: An increase in the number of *CSF3* -1719T alleles was associated with protection against low lung function, odds ratio (OR)=0.73, 95% confidence interval (CI)=0.56 to 0.95, P=0.018 and was still significant after adjustment for multiple comparisons. There was also a significant association of a *CSF3* haplotype with baseline FEV_1 levels (global test p=0.004 and 0.027 before and after adjustment for confounding factors). No association was found for *CSF2* SNPs and lung function, nor was there evidence of epistasis.

Conclusion: Genetic variation in *CSF3* is associated with cross-sectionally measured lung function in smokers.

INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is a complex genetic/environmental disorder that is characterized by airflow obstruction which is not fully reversible and 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 account 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 (GM-CSF), also known as colony-stimulating factor 2 (CSF2), is an important survival, proliferation and differentiation factor of the progenitor cells for neutrophils and macrophages. Granulocyte colony-stimulating factor (G-CSF), also known as CSF3, is specific for granulocytes. The *CSF2* gene (located at 5q31.1) and *CSF3* gene (located at 17q11.2-q12) were selected as candidates for decline and cross-sectional level of lung function in COPD patients based on following reasons. Firstly, CSF2 and CSF3 can induce the expression of pro-inflammatory cytokines and thereby enhance the inflammatory response. It was shown that CSF2 serum and bronchoalveolar lavage (BAL) levels, along with numbers of total cells and polymorphonuclear cells in the BAL were increased in bronchitic 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 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 these patients [14].

We hypothesized that *CSF2* and *CSF3* polymorphisms and their interactions will influence the decline of FEV_1 and/or the cross-sectional level of FEV_1 in smokers with mild to moderate air flow obstruction from the Lung Health Study (LHS) cohort. The LHS, sponsored by the US National Heart, Lung and Blood Institute, was a clinical trial of smoking intervention and bronchodilator treatment on the progression of COPD [15]. This 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 FEV_1% predicted as has been used previously [16-22].

METHODS

Study subjects:

The LHS recruited a total of 5887 smokers aged 35-60 with spirometric evidence of mild-moderate lung function impairment from 10 North American medical centers. Two nested case control studies were designed from the LHS cohort to study genetic determinants of rate of FEV_1 decline and cross-sectional level of FEV_1. Based on rate of decline of FEV_1 during 5 years of follow up, using arbitrary cut-off points of FEV_1 % predicted decrease $\geq$ 3.0%/year and increase $\geq$ 0.4%/year for rapid decliners and non decliners, respectively, we selected the 287 and the 308 non-Hispanic whites with the highest rate of decline of lung function (fast decline group) and the slowest rate of decline of lung function (non decline group) among 3,216 continuous smokers during the first 5 years of follow up. The rationale to select approximately the 300 highest and 300 lowest phenotypic subjects was that 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 Variants hypothesis was suggested one decade ago which 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 from our previous power analysis [28]. From all remaining LHS subjects, we selected non-Hispanic whites with the highest post bronchodilator FEV_1% predicted (high function group, n = 484) and the lowest post bronchodilator FEV_1% predicted (low function group, n = 468) at the beginning of the LHS. Arbitrary cut-off points of FEV_1 % predicted $\geq$ 88.9% and $\leq$ 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 limits which defined the cross-sectional groups (58 individuals were in the high function group and 86 individuals were in

the low lung function group), they were also analyzed in the study of cross sectional FEV_1. Thus, there were 542 and 554 subjects in the high and low lung function groups, respectively. Informed consent was obtained from all participants and this investigation received the approval of the Providence Health Care Research Ethics Board.

TagSNP selection:

The *CSF2* and *CSF3* SNP discovery data were downloaded from the SeattleSNPs NHLBI Program for Genomic Applications (PGA), UW-FHCRC, Seattle, WA (URL: http://pga.gs.washington.edu) [accessed October 2003]. From all SNPs identified in the 23 unrelated European-American samples from the Centre d'Etude Polymorphisme Humain family panel (CEPH), a set of tagSNPs was chosen for each gene using the LDSelect program developed by Carlson et al [29]. A LD threshold of $r^2 > 0.64$ and minor allele frequency of 5% were used. Two SNPs located at -1440A/G and 1944T/C (I117T) in the *CSF2* gene were selected initially, however the assay for the 1944T/C SNP could not be established by the TaqMan assay (ABI) and a PCR-RLFP assay for the same SNP showed that PCR amplification failed for some samples. Therefore, we replaced 1944T/C with an alternative SNP, 1622C/T. Three SNPs located at -1719C/T, -882G/A and 2176T/C in the *CSF3* gene 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 ABI (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 probes were labeled with 5' FAM or 5' VIC fluorophores as reporters (Applied Biosystems). Up to 47 DNA samples

of the CEPH panel with sequencing information available from the SeattleSNPs PGA were included as quality controls for each SNP genotyping. All genotype results from TaqMan assay were consistent with sequencing results for all CEPH DNA samples which 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 a Restriction Fragment Length Polymorphism-PCR (RFLP-PCR) method using the following primers flanking the polymorphic region: 5'-AAG GAA GGG AGG CTA CTT GG-3' (sense) and 5' GTT CCC CAA GGA GTG CAT AG-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 3 different genotypes. Sequencing was performed on an ABI 3100 16-capillary automated genetic analyzer (Applied Biosystems) using the same primers as in the PCR reaction to obtain the PCR product to be sequenced.

Statistical Analysis:

Hardy-Weinberg equilibrium tests and linkage disequilibrium estimation were done using the genetics package for R (www.r-project.org). All single-locus association tests were performed in R. The codominant and additive models were tested first, and if there was a significant association the dominant and recessive models were further tested to see if those models fit 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 a SNP in the case and control groups constitute a 2x3 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 [30]. 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 [30]. In a recessive model two copies of 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 additive model, there is r fold increased disease risk for heterozygotes compared with the homozygotes for the major allele, there is 2r fold increased disease risk for the homozygotes for the minor allele compared with the homozygotes for the major allele [30]. The Armitage trend test [31] was used to test an additive effect of the allele. In both the FEV_1 decline study and the cross-sectional FEV_1 study, in addition to crude analysis by chi-square tests using 2x3 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 FEV_1 level. In the FEV_1 decline study, multivariate logistic regression was used to adjust for confounding factors such as age, sex, pack-years of smoking, and research centre. In the cross-sectional FEV_1 study, multivariate logistic regression was used to adjust for the abovementioned confounding factors and rate of decline of FEV_1. Although other phenotypes such as FVC % predicted and FEV_1/FVC ratio were not our primary phenotypes due to our study design, we also analyzed associations of those phenotypes with single SNPs by 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 ≥ 5% (n_e was calculated by the equation, $n_e = \frac{1}{\sum_i p_i^2}$, where p_i is the frequency of the ith haplotype [29]. The

effective number 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 done on the basis of the spectral decomposition (SpD) of matrices of pairwise LD between SNPs by using SNPSpD (http://genepi.qimr.edu.au/general/daleN/SNPSpD/) [32]. This method provides a useful alternative to the very conservative Bonferroni correction. Haplotype association was tested using the hapassoc package for R. This software performs likelihood inference of trait associations with haplotypes and other covariates for generalized linear models, including logistic regression and does not assume haplotype phase is known [33]. 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.

Gene-gene interactions: A new approach "focused interaction testing framework" (FITF) was used to identify gene-gene interactions [34].

Power analysis:

The power of the two studies was estimated using the two independent proportions and many proportions functions in PASS 2005 (Hintze J, 2004. NCSS and PASS. Number Cruncher Statistical Systems. Kaysville, Utah. www.ncss.com). Plots were created in R (www.r-project.org) using the output from PASS.

RESULTS

I. Characteristics of the study groups:

The characteristics of study participants are shown in Tables 3 and 4. Because there was no DNA available for 8 subjects in the rate of decline of FEV_1 study and no DNA available for 22 subjects in the cross-sectional level of FEV_1 study, the numbers of participants in the two studies were 587 and 1074, respectively.

Among non decliners of the rate of decline of FEV_1 study and among the high lung function group of the cross-sectional FEV_1 study, the allele frequencies of all 5 SNPs did not significantly deviate from Hardy-Weinberg equilibrium (results not shown).

II. Haplotypes resolved with the genotyped tagSNPs:

Haplotypes from SNPs with minor allele frequency ≥ 5% in 23 CEPH samples were inferred by use of PHASE 2.0 [35, 36]. The LD-selected *CSF2* tagSNPs can resolve 60% (3 out of 5) of the actual number of haplotypes (Figure 1) and resolve 87.1% (2.7 out of 3.1) of the effective number (n_e) of haplotypes from SNPs with minor allele frequency ≥ 5%. For *CSF3*, 35.7% of actual haplotypes and 48.1% of effective haplotypes from SNPs with a minor allele frequency ≥ 5% were resolved by the 3 selected tagSNPs.

III. Single SNP association analysis:

In the FEV_1 decline study, none of the 5 SNPs were associated with decline of FEV_1 in co-dominant and additive models both before and after adjustment for confounding factors (Table 5).

In the cross-sectional level of FEV_1 study, 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 with p = 0.018 (Table 5). The Odds Ratio (OR) of having one -1719T allele compared with no -1719T allele and the OR of having two -1719T alleles compared with one -1719T allele was the same and was 0.73, 95% confidence interval (CI) 0.56 to 0.95. The association of *CSF3*_-1719 with FEV_1 level was adjusted for multiple testing on the basis of the SNP spectral decomposition approach [32]. The significance threshold required to keep Type I error rate at 5% for CSF3 in our study is 0.019 based on the LD of the 3 SNPs we 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 co-dominant model (comparison of the distribution of the three genotypic groups CC, CT and TT in the case and control groups, 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 > 0.1 for both SNPs (see Table 5).

Although FVC % predicted and FEV_1/FVC ratio at the beginning of the LHS were not our primary phenotypes as a result of our case-control study design, we performed exploratory analyses of single SNP associations with those phenotypes. In the rate of decline group, the FVC% predicted phenotype was normally distributed and therefore a one-way ANOVA was used to compare if FVC % predicted and FEV_1/FVC ratio were the same among the three genotypic groups. A significant association of FVC % predicted with CSF3_2176 was found

with P = 0.033 (Table 6), those individuals with the 2176TT genotype had a lower FVC % predicted. No other significant associations were found (data not shown).

IV. Haplotype association analysis:

Haplotypes from *CSF2* or *CSF3* were not associated with decline of FEV_1 in the analysis both without and with 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 7. The haplotypes from *CSF2* were not associated with decline of FEV_1 in the analysis both without and with adjustment for confounding factors. The three locus *CSF3* haplotypes were associated with levels of FEV_1 in a Wald global test (an overall test of haplotype distribution between cases and controls, p = 0.004) before adjustment for confounding factors, 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 versus low FEV1 group (16.9% versus 14.0%) when compared with the haplotype -1719C/-882G/2176T as a reference (adjusted p = 0.047). Analysis of two locus haplotypes (see Table 7) demonstrated that this marginal association was likely driven by both the -1719T allele and 2176C allele. The frequency of the haplotype -1719C/-882A/2176C was lower in the high versus low FEV1 group (34.2% versus 38.7%) when compared with the haplotype -1719C/-882G/2176T as a reference, but the significance became borderline when adjusting for confounding factors (unadjusted P = 0.007, adjusted p = 0.089).

V: Gene-gene interactions:

We explored *CSF2* and *CSF3* interaction of all possible two- to four-locus models using the "focused interaction testing framework" (FITF) method. There was no evidence of epistasis (gene-gene interaction) (detailed results not shown).

VI. Power of the study:

First, we calculated the power of our study for a codominant mode of inheritance. A chi-square test with 2 degrees of freedom was used to calculate the associated power. Effect size (a measure of the magnitude of the Chi-Square that is to be detected), 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 FEV_1 decline and cross-sectional FEV_1 level studies. The power of the dominant and recessive models was tested with a 2x2 table; the proportions in the control group were set to be close to those observed with the 5 SNPs in the "low" outcome groups (i.e. non decline of FEV_1 group and high lung function group). Figure 2 and Figure 3 give the curves of power versus OR value for the 5 studied SNPs for the baseline FEV_1 study for dominant and recessive models, respectively. For the FEV_1 decline study, the power was slightly less than that of the baseline FEV_1 study due to smaller sample size (Figures not shown).

DISCUSSION

CSF3 is a logical candidate gene for these studies due to its biologic function. In a rat model, neutrophil stimulation by CSF3 aggravates ventilator-induced lung injury manifested by increased lung neutrophils and IL-6 expression, increased alveolar edema on histology, and reduced lung compliance [37]. 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 as exon 4-165C>T in the original paper [13]) was associated with peripheral blood granulocyte count among workers exposed to benzene. 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 we did not find an association of CSF3_2176 with the primary phenotypes of baseline and decline of FEV_1, we found that a different SNP (*CSF3*_1719) was associated with baseline level of FEV_1. Interestingly, in an exploratory analysis of single SNPs with other phenotypes such as FVC % predicted and FEV_1/FVC ratio, a significant association of CSF3_2176 with FVC % predicted was found (without correction for multiple comparisons). The association of the CSF3_2176TT genotype with lower FVC % predicted 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 [38].

There are several explanations for the above observations including genetic heterogeneity between different populations, different phenotypes studied, and choice of tag SNPs. It was reported that tagSNPs selected using the criteria of R^2 of 0.64 and minor allele frequency of 5% could resolve 76% of actual and 85% of effective haplotypes in an analysis of 100 genes [29].

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 our population compared with that of the workers exposed to benzene [13], we might have missed the functional SNP in our study. The fact that our results showed that CSF3_1719 was associated with baseline level of FEV_1 and CSF3_2176 was associated with FVC % predicted suggests that neither SNP is causal but may be in linkage disequilibrium with a causal SNP which is yet to be identified.

There are several explanations for the observation that SNPs from *CSF3* but not *CSF2* were associated with lung function. Firstly, animal studies have documented that CSF3 plays a more important role than CSF2 in regulation of neutrophil homeostasis. Dogs depleted of CSF3 by a neutralizing antibody developed profound and selective neutropenia [39] but mice depleted of CSF2 did not show impairment of hematopoiesis [40]. In addition, *CSF3* but not *CSF2* knock-out mice display chronic neutropenia [41, 42]. 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 [43], 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 mobilizing granulocytes from the bone marrow, CSF2 and CSF3 are decisive in influencing the subsequent Th1 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 us to analyze gene-gene interaction. However, no significant CSF2 and CSF3 interaction was found in our study. There are several explanations for this: first, we might not have had enough power to detect gene-gene interaction with our sample size and minor allele frequencies. Second, cystic fibrosis with chronic *P. aeruginosa* lung infection is a Th2 dominated response [44] while COPD is a Th1 dominated response [45]. Therefore, the determinants of lung function in cystic fibrosis patients with chronic *P. aeruginosa* lung infection and in smoking induced COPD patients are likely different.

There are several concerns with this 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 such as ours [46, 47]. Second, 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, we only took into account multiple SNPs in a single gene. No correction for multiple genes and phenotypes was performed. Thirdly, we did not analyze a second cohort to replicate our results. Fourthly, no available function data support our associations. Finally, the 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 this study should be regarded as hypothesis generating only and it will be necessary to replicate them in different studies, especially in those with a cohort design.

In summary, we reported an association of the CSF3_-1719C/T with baseline level of FEV_1. However, this association needs to be replicated in different studies. Moreover, further functional study of this SNP or SNPs in LD with it is warranted.

Table 1. TagSNP selection using the LDSelect program and nomenclature of the SNPs.

Gene	Bin	SNP ID	SNP	Position in Ref 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)	-	-

Note: Sites are ordered by linkage disequilibrium, with sites showing similar patterns of genotype put into the same bin. The position in the gene is numbered by denoting the first nucleotide of the initiator methionine codon as +1 (position 1985 in sequence AF373868 of CSF2 and position 2040 in sequence AF388025 of CSF3). All SNPs with minor allele frequency > 5% are tagSNPs in this example. One SNP was genotyped in each bin and genotyped SNPs are indicated.

Table 2. *Taq*Man Primer and Probe Sequences.

SNP	Primer	Allelic Probe[a]
CSF2 -1440	Forward: AACTCCCACAGTACAGGGAAACTG Reverse: CAGAGAGCAGGTGGAGTTCATG	G: 6FAM ACTCAG**G**CCACAGTG MGBNFQ A: VIC CTCAG**A**CCACAGTGC MGBNFQ
CSF2 1622	Forward: GGGAAGGGAGCAAAGTTTGTG Reverse: AAACGCCTGCCTTTTTGGT	T[b]: 6FAM CAGCTG**A**GCTGAGG MGBNFQ C[b]: VIC AGCTG**G**GCTGAGGT MGBNFQ
CSF3 -1719	Forward: GCAATGAGCGAAACTCCATCTC Reverse: TGATGTGGCCCAGCTCTGTAC	C: VIC CCCAC**C**CTCTACTC MGBNFQ T: 6FAM CCCAC**T**CTCTACTCC MGBNFQ
CSF3 -882	Forward: CAGCCCGTGTCCACTTCAA Reverse: TTGGAACTGCGGGATTGG	G[b]: 6FAM ACGTGA**C**TTCCCTGGT MGBNFQ A[b]: VIC ACACGTGA**T**TTCC MGBNFQ
CSF3 2176	Forward: CAGGTGCCTGGACATTTGC Reverse: GTCTGCTCCCTCCCACATC	C[b]: 6FAM CAGTCCCC**G**TCCAGC MGBNFQ T[b]: VIC CAGTCCCC**A**TCCAGC MGBNFQ

[a] SNP sequences in the probes are highlighted in bold. [b] Probes are designed to the reverse strand.

Table 3. The distribution of demographic characteristics for the longitudinal FEV_1 change study

	Fast Decliners (n = 281)	Non Decliners (n = 306)	p value
Male/Female	164/117	204/102	0.038
Age (years)	49.51 ± 0.38	47.61 ± 0.39	0.0006
Smoking history (pack-yrs)[a]	42.86 ± 1.14	38.38 ± 1.04	0.004
ΔFEV_1/yr (% predicted pre)[b]	-4.14 ± 0.06	1.08 ± 0.04	<0.0001
Baseline FEV_1 (% predicted pre)[c]	72.6 ± 0.53	75.7 ± 0.46	<0.0001

Values are mean ± SE for continuous data. [a] Number of packs of cigarettes smoked per day × number of years smoking. [b] Change in FEV_1 over a 5 year period per year as % predicted FEV_1 pre bronchodilator. [c] FEV_1 at the start of the LHS as measured FEV_1(%) predicted pre bronchodilator

Table 4. The distribution of demographic characteristics for the cross-sectional FEV_1 study

	High FEV_1 (n = 536)	Low F EV_1 (n = 538)	p value
Male/Female	354/182	332/206	0.139
Age (years)	46.24 ± 0.30	50.69 ± 0.26	<0.0001
Smoking history (pack-yrs)	35.32 ± 0.77	45.16 ± 0.81	<0.0001
ΔFEV_1/yr (% predicted pre)[a]	-0.55 ± 0.07	-1.27 ± 0.08	<0.0001
ΔFEV_1/yr (% predicted post)[b]	-0.75 ± 0.06	-0.79 ± 0.08	<0.722
Baseline FEV_1 (% predicted pre)[a]	86.48 ± 0.13	61.08 ± 0.18	<0.0001
Baseline FEV_1 (% predicted post)[b]	91.80 ± 0.10	62.61 ± 0.14	<0.0001

Values are mean ± SE for continuous data. [a] prebronchodilator. [b] postbronchodilator.

Table 5. Single SNP association of CSF2 and CSF3 genes in FEV_1 longitudinal decline and cross-sectional level of FEV_1 studies

SNP	Genotype[a]	Rapid Decliners	Non-decliners	Co-dominant[b]		Additive[c]		Low function	High function	Co-dominant		Additive	
		n (%)	n (%)	P[d]	P[e]	P[d]	P[e]	n (%)	n (%)	P[d]	P[f]	P[d]	P[f]
CSF2 -1440	AA	100 (35.7)	101 (33.0)					166 (32.2)	180 (34.2)				
	AG	124 (44.3)	152 (49.7)	n.s.	n.s.	n.s.	n.s.	252 (48.9)	259 (49.2)	n.s.	n.s.	n.s.	n.s.
	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)	n.s.	n.s.	n.s.	n.s.	140 (27.0)	171 (32.4)	0.092	n.s.	n.s.	n.s.
	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)	n.s.	n.s.	n.s.	n.s.	124 (23.8)	147 (28.0)	n.s.	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)	n.s.	n.s.	n.s.	n.s.	234 (45.3)	233 (45.0)	n.s.	n.s.	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)	n.s.	n.s.	n.s.	n.s.	247 (47.3)	259 (49.3)	n.s.	n.s.	n.s.	n.s.
	TT	47 (17.2)	37 (12.3)					83 (15.9)	80 (15.3)				

[a] Homozyotes for the major allele are listed first.
[b] Codominant model: the three genotypes of a SNP in the cases and controls constitute a 2x3 contingency table, the direct calculation of 2x3 tables provides unadjusted p values, while multiple regression models were used to adjust for confounding factors.

[c] additive model: there is r fold increased disease risk for heterozygotes compared with the homozygotes for the major allele, there is 2r fold increased disease risk for the homozygotes for the minor allele compared with the homozygotes for the major allele. The Armitage trend test was used to test an additive effect of the allele.
[d] Unadjusted P value
[e] Adjusted for age, sex, center, smoking history (pack years)
[f] Adjusted for age, sex, center, smoking history (pack years) and rate of decline of lung function (% predicted postbronchodilator)

Table 6. Single SNP association of *CSF2* and *CSF3* with FVC % predicted in the longitudinal decline in FEV_1 study group (overall comparison of differences of FVC % among three genotype groups)

SNP	N	Genotype	FVC % predicted (mean ± SE)	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		

Table 7. Haplotype association of CSF2 and CSF3 in the cross-sectional level of FEV_1 study.

Gene	Haplotype[a]	Low function	High function	Global test P value[b]		P value[c]	
		%	%	P[d]	P[e]	P[d]	P[e]
CSF2	-1440A/1622C	33.0	36.9				
	-1440G/1622C	49.7	44.3	n.s.	n.s.	n.s.	n.s.
	-1440G/1622T	17.3	20.0			n.s.	n.s.
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			n.s.	0.047
	-1719C/-882G/2176C	7.5	8.9			n.s.	n.s.
CSF3	-1719C/-882G	48.4	48.8				
	-1719C/-882A	38.7	34.2	0.069	0.0427	n.s.	n.s.
	-1719T/-882G	14.0	16.9			n.s.	0.058
	-882G/2176T	39.9	39.9				
	-882A/2176C	38.7	34.2	0.051	0.053	n.s.	n.s.
	-882G/2176C	21.5	25.8			n.s.	0.081

[a] Haplotype frequencies were calculated using an Expectation Maximization algorithm from the haplo.stats package for R

[b] comparing overall haplotype distribution between cases and controls.

[C] comparing each haplotype between cases and controls using the most common haplotype as a reference.

[d] Unadjusted P value

[e] Adjusted for age, sex, center, smoking history (pack years) and rate of decline of lung function (% predicted postbronchodilator)

REFERENCES

1. Astemborski JA, Beaty TH, Cohen BH. Variance components analysis of forced expiration in families. *American Journal of Medical Genetics* 1985: 21(4): 741-753.
2. Lewitter FI, Tager IB, McGue M, Tishler PV, Speizer FE. Genetic and environmental determinants of level of pulmonary function. *American Journal of Epidemiology* 1984: 120(4): 518-530.
3. 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(1): 81-94.
4. Gottlieb DJ, Wilk JB, Harmon M, Evans JC, Joost O, Levy D, O'Connor GT, Myers RH. Heritability of longitudinal change in lung function. The Framingham study. *Am J Respir Crit Care Med* 2001: 164(9): 1655-1659.
5. Balbi B, Bason C, Balleari E, Fiasella F, Pesci A, Ghio R, Fabiano F. Increased bronchoalveolar granulocytes and granulocyte/macrophage colony-stimulating factor during exacerbations of chronic bronchitis. *Eur Respir J* 1997: 10(4): 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(5): 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(1): 247-248.

8. Rafatpanah H, Bennett E, Pravica V, McCoy MJ, David TJ, Hutchinson IV, Arkwright PD. Association between novel GM-CSF gene polymorphisms and the frequency and severity of atopic dermatitis. *J Allergy Clin Immunol* 2003: 112(3): 593-598.

9. He JQ, Ruan J, Chan-Yeung M, Becker AB, Dimich-Ward H, Pare PD, Sandford AJ. Polymorphisms of the GM-CSF genes and the development of atopic diseases in at-risk children. *Chest* 2003: 123(3 Suppl): 438S.

10. Sluiter HJ, Koeter GH, de Monchy JG, Postma DS, de Vries K, Orie NG. The Dutch hypothesis (chronic non-specific lung disease) revisited. *European Respiratory Journal* 1991: 4(4): 479-489.

11. Vestbo J, Prescott E. Update on the "Dutch hypothesis" for chronic respiratory disease. *Thorax* 1998: 53(Suppl 2): S15-19.

12. Weiss ST. Atopy as a risk factor for chronic obstructive pulmonary disease: epidemiological evidence. *Am J Respir Crit Care Med* 2000: 162(3 Pt 2): S134-136.

13. Lan Q, Zhang L, Shen M, Smith MT, Li G, Vermeulen R, Rappaport SM, Forrest MS, Hayes RB, Linet M, Dosemeci M, Alter BP, Weinberg RS, Yin S, Yeager M, Welch R, Waidyanatha S, Kim S, Chanock S, Rothman N. Polymorphisms in cytokine and cellular adhesion molecule genes and susceptibility to hematotoxicity among workers exposed to benzene. *Cancer Res* 2005: 65(20): 9574-9581.

14. Moser C, Jensen PO, Pressler T, Frederiksen B, Lanng S, Kharazmi A, Koch C, Hoiby N. 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(6): 400-409.

15. Anthonisen NR, Connett JE, Kiley JP, Altose MD, Bailey WC, Buist AS, Conway WA, Enright PL, Kanner RE, O' Hara P, Owens GR, Scanlon PD, Tashkin DP, Wise RA. Effects of smoking intervention and the use of an inhaled anticholinergic bronchodilator on the rate of decline of FEV_1. The Lung Health Study. *JAMA* 1994: 272(19): 1497-1505.

16. He JQ, Connett JE, Anthonisen NR, Pare PD, Sandford AJ. Glutathione S-transferase variants and their interaction with smoking on lung function. *Am J Respir Crit Care Med* 2004: 170(4): 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(3): 379-385.

18. He JQ, Ruan J, Connett J, Anthonisen N, Paré P, Sandford A. Antioxidant gene polymorphisms and susceptibility to a rapid decline in lung function in smokers. *Am J Respir Crit Care Med* 2002: 166(3): 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(2): 469-473.

20. Joos L, He JQ, Shepherdson MB, Connett JE, Anthonisen NR, Pare PD, Sandford AJ. The role of matrix metalloproteinase polymorphisms in the rate of decline in lung function. *Hum Mol Genet* 2002: 11(5): 569-576.

21. He JQ, Burkett K, Connett JE, Anthonisen NR, Pare PD, Sandford AJ. Interferon gamma polymorphisms and their interaction with smoking are associated with lung function. *Hum Genet* 2006: 119(4): 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(6): 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. *Human Heredity* 1998: 48(2): 97-107.

24. Ernster VL. Nested case-control studies. *Prev Med* 1994: 23(5): 587-590.

25. Lander ES. The new genomics: global views of biology. *Science* 1996: 274(5287): 536-539.

26. Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N, Shaw N, Lane CR, Lim EP, Kalyanaraman N, Nemesh J, Ziaugra L, Friedland L, Rolfe A, Warrington J, Lipshutz R, Daley GQ, Lander ES. Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nature Genetics* 1999: 22(3): 231-238.

27. Chakravarti A. Population genetics--making sense out of sequence. *Nat Genet* 1999: 21(1 Suppl): 56-60.

28. Tanaka G, Sandford AJ, Burkett K, Connett JE, Anthonisen NR, Pare PD, He JQ. Tumour necrosis factor and lymphotoxin A polymorphisms and lung function in smokers. *Eur Respir J* 2006: 29(1): 34-41.

29. 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(1): 106-120.

30. Lewis CM. Genetic association studies: design, analysis and interpretation. *Brief Bioinform* 2002: 3(2): 146-153.

31. Armitage P. Tests for Linear Trends in Proportions and Frequencies. *Biometrics* 1955: 11(3): 375-386.

32. Nyholt DR. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am J Hum Genet* 2004: 74(4): 765-769. Epub 2004 Mar 2002.

33. 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(4): 200-206.

34. 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(1): 15-27.

35. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 2001: 68(4): 978-989.

36. Stephens M, Donnelly P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet* 2003: 73(5): 1162-1169.

37. Karzai W, Cui X, Heinicke N, Niemann C, Gerstenberger EP, Correa R, Banks S, Mehlhorn B, Bloos F, Reinhart K, Eichacker PQ. Neutrophil stimulation with granulocyte colony-stimulating factor worsens ventilator-induced lung injury and mortality in rats. *Anesthesiology* 2005: 103(5): 996-1005.

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

39. Hammond WP, Csiba E, Canin A, Hockman H, Souza LM, Layton JE, Dale DC. Chronic neutropenia. A new canine model induced by human granulocyte colony-stimulating factor. *J Clin Invest* 1991: 87(2): 704-710.

40. Dranoff G, Crawford AD, Sadelain M, Ream B, Rashid A, Bronson RT, Dickersin GR, Bachurski CJ, Mark EL, Whitsett JA, et al. Involvement of granulocyte-macrophage colony-stimulating factor in pulmonary homeostasis. *Science* 1994: 264(5159): 713-716.

41. Stanley E, Lieschke GJ, Grail D, Metcalf D, Hodgson G, Gall JA, Maher DW, Cebon J, Sinickas V, Dunn AR. Granulocyte/macrophage colony-stimulating factor-deficient mice show no major perturbation of hematopoiesis but develop a characteristic pulmonary pathology. *Proc Natl Acad Sci U S A* 1994: 91(12): 5592-5596.

42. Lieschke GJ, Grail D, Hodgson G, Metcalf D, Stanley E, Cheers C, Fowler KJ, Basu S, Zhan YF, Dunn AR. Mice lacking granulocyte colony-stimulating factor have chronic neutropenia, granulocyte and macrophage progenitor cell deficiency, and impaired neutrophil mobilization. *Blood* 1994: 84(6): 1737-1746.

43. 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(5): 2663-2669.

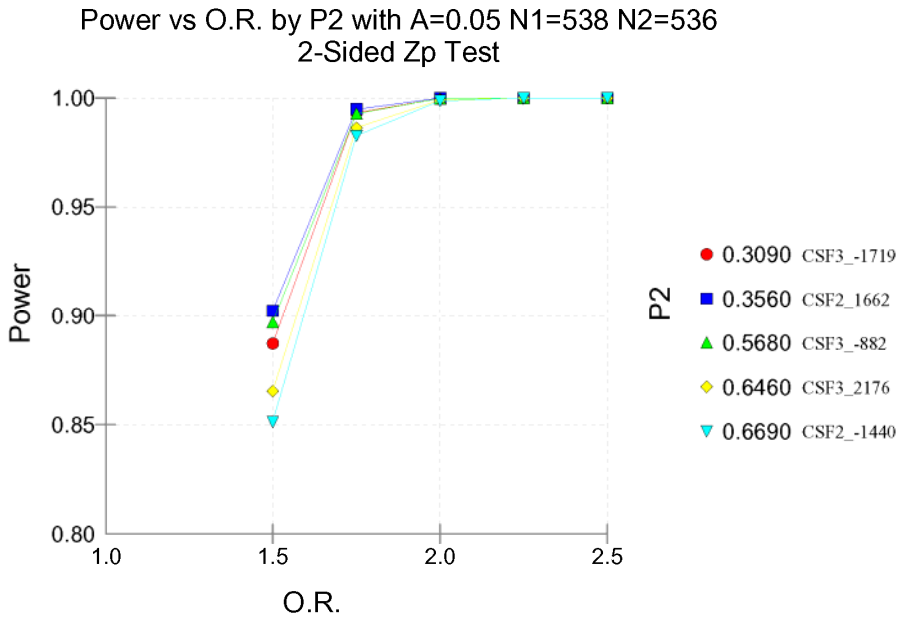
44. Moser C, Kjaergaard S, Pressler T, Kharazmi A, Koch C, Hoiby N. The immune response to chronic Pseudomonas aeruginosa lung infection in cystic fibrosis patients is predominantly of the Th2 type. *Apmis* 2000: 108(5): 329-335.

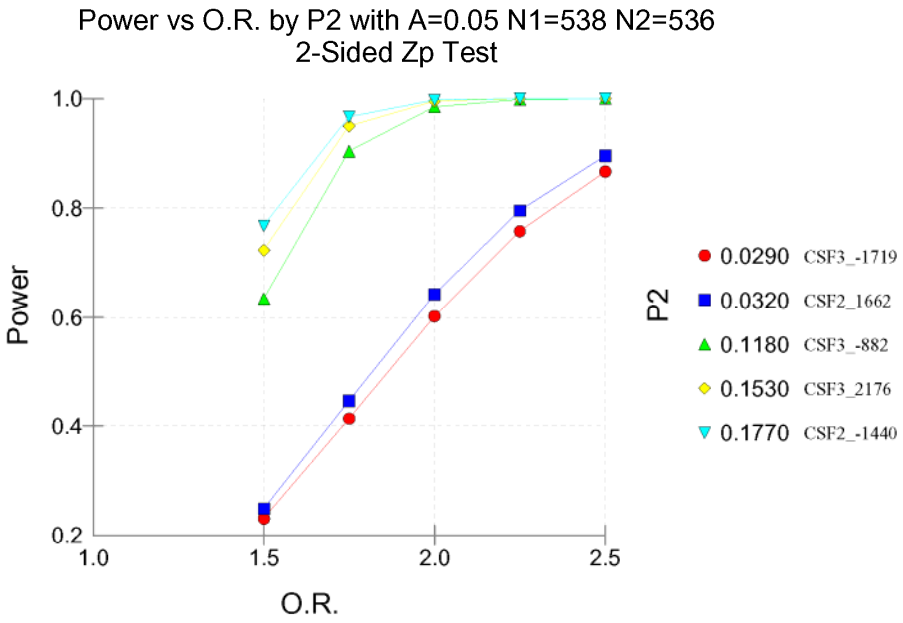
45. Majori M, Corradi M, Caminati A, Cacciani G, Bertacco S, Pesci A. Predominant TH1 cytokine pattern in peripheral blood from subjects with chronic obstructive pulmonary disease. *J Allergy Clin Immunol* 1999: 103(3 Pt 1): 458-462.

46. Ardlie KG, Lunetta KL, Seielstad M. Testing for population subdivision and association in four case-control studies. *Am J Hum Genet* 2002: 71(2): 304-311.

47. 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(14): 1151-1158.

Haplotype	Frequency (%)	SNP position in the CSF2 gene (Genotyped tagSNPs are denoted in bold and underlined)								
		-1916	**-1440**	-675	1363	**1622**	1944	2416	2722	3739
1	41.30									
2	34.78									
3	17.39									
4	4.35									
5	2.17									


Power vs O.R. by P2 with A=0.05 N1=538 N2=536
2-Sided Zp Test
Power
1.00
0.95
0.90
0.85
0.80
1.0
1.5
2.0
2.5
O.R.
P2
0.3090 CSF3_-1719
0.3560 CSF2_1662
0.5680 CSF3_-882
0.6460 CSF3_2176
0.6690 CSF2_-1440


Power vs O.R. by P2 with A=0.05 N1=538 N2=536
2-Sided Zp Test
Power
1.0
0.8
0.6
0.4
0.2
1.0
1.5
2.0
2.5
O.R.
P2
0.0290 CSF3_-1719
0.0320 CSF2_1662
0.1180 CSF3_-882
0.1530 CSF3_2176
0.1770 CSF2_-1440