



Candidate genes for COPD in two large data sets

P.S. Bakke*, G. Zhu*, A. Gulsvik, X. Kong, A.G.N. Agusti, P.M.A. Calverley, C.F. Donner, R.D. Levy, B.J. Make, P.D. Paré, S.I. Rennard, J. Vestbo, E.F.M. Wouters, W. Anderson, D.A. Lomas, E.K. Silverman and S.G. Pillai

ABSTRACT: Lack of reproducibility of findings has been a criticism of genetic association studies on complex diseases, such as chronic obstructive pulmonary disease (COPD).

We selected 257 polymorphisms of 16 genes with reported or potential relationships to COPD and genotyped these variants in a case–control study that included 953 COPD cases and 956 control subjects. We explored the association of these polymorphisms to three COPD phenotypes: a COPD binary phenotype and two quantitative traits (post-bronchodilator forced expiratory volume in 1 s (FEV₁) % predicted and FEV₁/forced vital capacity (FVC)). The polymorphisms significantly associated to these phenotypes in this first study were tested in a second, family-based study that included 635 pedigrees with 1,910 individuals.

Significant associations to the binary COPD phenotype in both populations were seen for *STAT1* (rs13010343) and *NFKB1B/SIRT2* (rs2241704) ($p < 0.05$). Single-nucleotide polymorphisms rs17467825 and rs1155563 of the *GC* gene were significantly associated with FEV₁ % predicted and FEV₁/FVC, respectively, in both populations ($p < 0.05$).

This study has replicated associations to COPD phenotypes in the *STAT1*, *NFKB1B/SIRT2* and *GC* genes in two independent populations, the associations of the former two genes representing novel findings.

KEYWORDS: Chronic obstructive pulmonary disease, genetic association, replication, smoking

The prevalence of chronic obstructive pulmonary disease (COPD) in Western Europe is ~10% [1], and COPD is expected to be the third most significant cause of death worldwide by 2020 [2]. The most important risk factor for COPD is smoking and there is a dose–response relationship between smoking exposure and reduced lung function, although there is a substantial heterogeneity in the degree of lung function impairment [3]. Only a subset of smokers develops an accelerated rate of decline in lung function that leads to COPD. In addition, there appears to be familial clustering of both impaired lung function and COPD [4]. These insights suggest that susceptibility to COPD may be influenced by genetic factors. The only well-established genetic cause of COPD, α_1 -antitrypsin deficiency, is present in only 1–2% of individuals with COPD [5]. A number of studies have been performed to find other genetic susceptibility factors for COPD. So far, hundreds of candidate genes have been tested. To date, it has been difficult to replicate genetic findings from one

COPD study to another [6]. There may be several explanations for this lack of reproducibility, including small sample sizes, lack of Hardy–Weinberg equilibrium (HWE), poor phenotype characterisation of the COPD cases and genetic heterogeneity [7]. It is now recommended that the findings of a genetic association should be replicated in another sample before being published [7]; nevertheless, only a minority of genetic COPD studies meets this requirement.

To obtain further insight into the genetic basis of COPD, we replicated the relationships of a number of potential COPD candidate genes in two large, independent and well-characterised populations. We selected 257 single-nucleotide polymorphisms (SNPs) in 16 genes based on reported or potential relationships to COPD. They were analysed in a case–control sample from Bergen, Norway, including 953 COPD cases and 956 controls [8]. SNPs with significant associations to COPD were then tested using family-based association analysis in 635 pedigrees with 1,910

AFFILIATIONS

A full list of the authors' affiliations can be found in the Acknowledgements section.

*These authors contributed equally to the study.

CORRESPONDENCE

P.S. Bakke

Dept of Thoracic Medicine
Haukeland University Hospital
N-5021 Bergen
Norway

E-mail: per.bakke@med.uib.no

Received:

June 10 2009

Accepted after revision:

June 04 2010

First published online:

June 18 2010

This article has supplementary material available from www.erj.ersjournals.com

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

individuals from the International COPD Genetics Network (ICGN), which is the largest family-based COPD collection reported to date [8].

METHODS

Study subjects

The Norwegian case-control study initially included 953 cases; 189 were recruited from two community studies [9], while the rest were recruited from a registry at Haukeland University Hospital, Bergen. The study also included 956 controls; 735 were recruited from the two community studies, while 221 were volunteers. The inclusion criteria for COPD cases was a post-bronchodilator forced expiratory volume in 1 s (FEV₁) <80% predicted and FEV₁/forced vital capacity (FVC) <0.7. The controls were selected based on post-bronchodilator FEV₁ >80% pred and FEV₁/FVC >0.7. Both cases and controls were Caucasians with ≥2.5 pack-yrs smoking history (current or ex-smokers).

In the multicentre ICGN study, subjects with known COPD were recruited as probands, and siblings and available parents were ascertained through the probands [8, 10]. The probands were recruited from pulmonary and medical clinics, and hospital admissions. Inclusion criteria for probands were airflow limitation (post-bronchodilator FEV₁ <60% pred and FEV₁/vital capacity (VC) <90% pred) at a relatively early age (45–65 yrs), a ≥5 pack-yrs smoking history and at least one eligible sibling (with ≥5 pack-yrs smoking history). COPD in siblings was defined by a post-bronchodilator FEV₁ <80% pred and FEV₁/VC <90% pred. 1,910 Caucasian individuals from 635 pedigrees were included in the family-based association analysis.

Phenotyping

Three phenotypes were defined: 1) the binary COPD phenotype defined according to the criteria described previously; 2) post-bronchodilator FEV₁ % pred; and 3) FEV₁/(F)VC.

Candidate genes selection and genotyping

16 candidate genes were selected for analyses based on their potential biological relevance to pathways that may cause COPD or their proximity to genes with a known relationship to COPD (see supplementary table E1 for details of each gene).

Seven out of the 16 genes had previously been shown to be associated with COPD (*GC*, *GSTP1*, *HDAC2*, *HDAC5*, *HMOX1*, *IL11* and *JAK3*). Genotyping in the two cohorts was performed with the Illumina (San Diego, CA, USA) array-based custom SNP genotyping platform. The selection of the candidate genes and the analyses were performed prior to the genome-wide association study (GWAS) that has recently been published using a subset of these subjects [11], which used an Illumina HumanHap550 genotyping BeadChip for the Bergen cohort and a Sequenom (Hamburg, Germany) iPLEX SNP genotyping protocol developed for measurement with the MassARRAY mass spectrometer for the ICGN study. The six SNPs rs13010343, rs1609181, rs802372, rs10278590, rs8065686 and rs4802898 were also present on the Illumina Human Hap550 chip from the GWAS [11].

For the selection of SNPs, linkage disequilibrium (LD) bins were established using an r^2 threshold of 0.8. The tagging SNP selection was based on HapMap data for European-Americans

with a minor-allele frequency (MAF) >5% from the public database [12]. Nonsynonymous SNPs with any MAF were included. SNPs in these genes were selected and genotyped in both the ICGN family population and the Norwegian case-control population using the Illumina array-based custom SNP genotyping platform. HWE was performed for all SNPs in the control data by using the Chi-squared goodness-of-fit test with SAS software version 8.2 (SAS Institute, Cary, NC, USA); HWE for all SNPs was also tested in the family data using PBAT version 3.5 (Golden Helix Inc., Bozeman, MT, USA) [9]. All SNPs (p-values >0.05) were in HWE in both the family data and the case-control data. COPD family data were evaluated for inconsistent Mendelian inheritance using the PedCheck program [13]. A complete list of the genes and SNPs tested in each gene is given in supplementary table E2.

Statistical methods

In the case-control population, two models were used in the association analysis. A logistic regression model for the COPD binary phenotype and a linear regression model for the quantitative phenotypes (FEV₁ and FEV₁/FVC), with covariates including age, sex and pack-yrs of smoking. For the quantitative trait analysis, only COPD cases were included. The analyses were done using SAS software version 8.2 with an additive genetic model. FBAT version 1.7.3 [14] was used for the family-based single-SNP association analysis of the COPD binary phenotype in the ICGN family study. The analyses of quantitative traits (FEV₁ and FEV₁/VC) were performed with covariates including centre, age, sex, height and pack-yrs of cigarette smoking using PBAT version 3.5 [14]. Biallelic tests were conducted for SNPs using an additive genetic model. The risk allele was determined from the FBAT Z statistic. Haplotype analyses were conducted using the HBAT function of the FBAT program with the use of Monte Carlo sampling for COPD, FEV₁ and FEV₁/VC [15] in the family data. In the case-control data, haplotype analysis was performed using the expectation-maximisation algorithm and score tests, implemented in the Haplo.stats program [16]. LD structure was examined with the Haploview program [17, 18]. We used a p-value <0.05 in both COPD populations to define statistical significance.

We assessed the power to detect significant associations between the genes and the phenotypes based on the following assumptions: we assumed that the allele frequencies of disease gene and the marker were 0.1, and the penetrances for the three genotypes of the marker were 0.7, 0.4 and 0.1. At significance level of 0.05, our study had 99.52 and 85.43% power for COPD case-control and family data, respectively, to detect an association.

RESULTS

Study participants

Characteristics of the participants of the two studies are given in table 1. In the case-control study, the cases comprised more males, were older and reported higher smoking exposure than the controls. The cases had, on average, a moderate-to-severe airflow limitation (mean post-bronchodilator FEV₁ 50.3% pred). In the ICGN study, the probands were predominantly male, of the same mean age, but with a greater smoking history than the siblings. The probands had, on average, a severe airflow limitation (mean post-bronchodilator FEV₁ 36.2% pred).

TABLE 1 Characteristics of the Norwegian case–control study and the International COPD Genetics Network (ICGN) study

	COPD case–control study		ICGN family study	
	Cases	Controls	Probands	Siblings
Subjects	953	956	610	1300
Age yrs	65.5 ± 10.1	55.5 ± 9.7	58.4 ± 5.5	58.1 ± 9.8
Females	370 (38.8)	477 (49.9)	247 (40.5)	648 (49.80)
Post-bronchodilator FEV ₁ L	1.58 ± 0.71	3.24 ± 0.74	1.11 ± 0.44	2.36 ± 0.98
Post-bronchodilator FEV ₁ % pred	50.3 ± 17.5	94.1 ± 9.2	36.3 ± 12.9	77.5 ± 25.9
Post-bronchodilator FEV ₁ /FVC [#]	0.51 ± 0.13	0.79 ± 0.04	0.37 ± 0.12	0.61 ± 0.15
Smoking exposure pack-yrs	32.0 ± 18.6	19.1 ± 13.2	51.5 ± 26.8	40.6 ± 24.9
Current smokers	457 (48.1)	407 (42.6)	207 (33.9)	661 (50.8)

Data are presented as n, mean ± SD or n (%). COPD: chronic obstructive pulmonary disease; FEV₁: forced expiratory volume in 1 s; % pred: % predicted; FVC: forced vital capacity. [#]: FEV₁/vital capacity (VC) was used for the ICGN family study, with VC determined as the higher of FVC and synchronised VC.

Single-marker and haplotype analysis

The relationships of the single-SNP associations with COPD, post-bronchodilator FEV₁ % pred and FEV₁/FVC (FEV₁/VC in the ICGN study) are shown in tables 2, 3 and 4, respectively. Only genes with significant associations in the case–control study are shown.

The following genes included SNPs that were significantly associated with the binary COPD phenotype in either study: *STAT1*, *GC*, *MAP3K5*, *KCND2*, *RARRES2*, *HADC5*, *SIRT2* and *NFKBIB*. Only one SNP in *STAT1* (rs13010343) and one SNP in *NFKBIB/SIRT2* (rs2241704) were associated with COPD in both studies (table 2). The risk allele was the same in both datasets for rs13010343, as well as for rs2241704.

The genes significantly associated with post-bronchodilator FEV₁ % pred in either study were *STAT1*, *GC*, *AGER*, *MAP3K5*, *KCND2*, *JAK3*, *SIRT2*, *NFKBIB/SIRT2*, *IL11* and *HMOX1*. Only *GC* (rs17467825) was replicated in both studies (table 3), with negative regression coefficients indicating a risk for impaired FEV₁.

Regarding the associations with FEV₁/FVC or FEV₁/VC, significant associations were noted for *STAT1*, *GC*, *AGER*, *HDAC2*, *MAP3K5*, *KCND2* and *JAK3* in either study. Only *GC* (rs1155563) was observed in both samples (table 4), with negative regression coefficients indicating a risk for impaired ratio.

No SNP had replicated associations across the two studies for all the COPD phenotypes.

We also performed haplotype analyses. The results supported the findings of the SNP analyses, but no additional significantly associated genes were identified (data not shown).

LD analysis

Figures 1–3 show the pair-wise LD (r^2) values for the SNPs in the genes *STAT1*, *NFKBIB/SIRT2* and *GC*. The LD structure appeared generally quite similar across the two study populations. Three haplotype blocks were identified in the LD map of the *STAT1* gene in both populations, with significant SNPs given in bold text. Two haplotype blocks

TABLE 2 Single-SNP association analysis with chronic obstructive pulmonary disease (COPD)

Gene	SNP	COPD case–control study			ICGN family study		
		Risk allele	OR (95% CI)	p-value	Risk allele	Informative families n	p-value
<i>STAT1</i>	rs12468579	G	1.18 (1.01–1.37)	0.0307	G	220	0.5505
<i>STAT1</i>	rs13010343	G	1.24 (1.01–1.55)	0.0483	G	113	0.0371
<i>KCND2</i>	rs1861064	G	0.79 (0.62–0.99)	0.0469	G	123	0.3622
<i>KCND2</i>	rs1609181	T	0.69 (0.51–0.95)	0.0225	T	61	0.4720
<i>KCND2</i>	rs802358	A	0.79 (0.63–0.99)	0.0432	A	123	0.5845
<i>KCND2</i>	rs802372	C	0.73 (0.54–0.97)	0.0311	C	77	0.8047
<i>KCND2</i>	rs4730979	G	1.22 (1.03–1.44)	0.0162	A	194	0.1882
<i>RARRES2</i>	rs10278590	G	1.17 (1.01–1.37)	0.0380	T	217	0.6587
<i>HDAC5</i>	rs8065686	C	1.48 (1.12–1.95)	0.0051	T	75	0.6582
<i>SIRT2</i>	rs4802998	A	1.17 (1.01–1.37)	0.0402	A	205	0.8746
<i>NFKBIB/ SIRT2</i>	rs2241704	T	0.81 (0.67–0.96)	0.0214	T	172	0.0442

Only single nucleotide polymorphisms (SNPs) with p-values < 0.05 in the COPD case–control study are shown. Bold represents significant replication in both populations. ICGN: International COPD Genetics Network.

TABLE 3 Single-SNP association analysis with forced expiratory volume in 1 s

Gene	SNP	COPD case-control study			ICGN family study		
		Risk allele	β (95% CI)	p-value	Risk allele	Informative families n	p-value
<i>NFE2L2</i>	rs2706110	G	-0.071 (-0.141– -0.002)	0.0438	G	228	0.5608
<i>STAT1</i>	rs3771300	C	-0.073 (-0.145– -0.002)	0.0445	C	148	0.3155
<i>STAT1</i>	rs13010343	A	0.113 (0.020–0.207)	0.0170	A	176	0.1692
<i>STAT1</i>	rs16833177	T	-0.070 (-0.138– -0.003)	0.0401	C	231	0.2745
GC	rs17467825	A	-0.067 (-0.127– -0.008)	0.0257	A	298	0.0462
<i>GC</i>	rs2282680	G	-0.069 (-0.129– -0.010)	0.0222	G	297	0.0717
<i>GC</i>	rs4588	C	-0.071 (-0.131– -0.012)	0.0187	C	298	0.0665
<i>GC</i>	rs1155563	T	-0.078 (-0.136– -0.021)	0.0078	T	300	0.1389
<i>AGER</i>	rs3131297	T	0.114 (0.048–0.179)	0.0006	T	259	0.6431
<i>AGER</i>	rs2269423	T	-0.066 (-0.118– -0.012)	0.0148	T	332	0.7342
<i>AGER</i>	rs204994	A	0.081 (0.014–0.147)	0.0173	A	272	0.3190
<i>AGER</i>	rs204993	T	0.084 (0.022–0.146)	0.0078	T	303	0.5362
<i>KCND2</i>	rs7793864	A	-0.104 (-0.205– -0.004)	0.0411	T	123	0.7888
<i>KCND2</i>	rs802340	T	-0.086 (-0.162– -0.010)	0.0263	G	221	0.8958
<i>SIRT2</i>	rs7257949	A	0.105 (0.002–0.209)	0.0463	A	117	0.4301
<i>SIRT2</i>	rs11879029	G	0.108 (0.002–0.214)	0.0460	G	118	0.3010
<i>SIRT2</i>	rs7251489	G	0.109 (0.004–0.214)	0.0416	G	116	0.3912
<i>SIRT2</i>	rs10410544	C	-0.063 (-0.118– -0.009)	0.0217	C	319	0.5244
<i>NFKBIB/SIRT2</i>	rs2053071	G	0.066 (0.012–0.122)	0.0175	G	311	0.9974
<i>IL11</i>	rs4252552	G	0.107 (-0.007–0.221)	0.0675	G	109	0.0489
<i>HMOX1</i>	rs17885925	C	-0.143 (-0.271– -0.015)	0.0281	C	76	0.3921
<i>HMOX1</i>	rs11912889	A	-0.147 (-0.274– -0.019)	0.0235	A	76	0.3283

Only single nucleotide polymorphisms (SNPs) with p-values <0.05 in the chronic obstructive pulmonary disease (COPD) case-control study are shown. Bold represents significant replication in both populations. ICGN: International COPD Genetics Network; β : regression coefficient.

and one haplotype block were revealed for the *NFKBIB/SIRT2* gene ICGN and case-control samples, respectively, while three haplotype blocks were observed in the two data sets for the *GC* gene. The six SNPs that were also present in the Illumina chip from the GWAS was not in significant LD ($r^2 \geq 0.8$) with any other SNPs tested in this study.

DISCUSSION

In the present study, we selected 257 polymorphisms in 16 genes with reported or potential relationships to COPD and genotyped the variants in a case-control COPD study. Significant associations were tested in a family-based COPD study. We detected that the *STAT1*, *NFKBIB/SIRT2* and *GC*

TABLE 4 Single-SNP association analysis with forced expiratory volume in 1 s (FEV₁)/forced vital capacity (FVC)[#]

Gene	SNP	COPD case-control study			ICGN family study		
		Risk allele	β (95% CI)	p-value	Risk allele	Informative families n	p-value
<i>STAT1</i>	rs13010343	A	0.024 (0.004–0.044)	0.0178	A	176	0.3393
GC	rs1155563	T	-0.014 (-0.0275– -0.002)	0.0203	T	300	0.0281
<i>AGER</i>	rs3131297	T	0.024 (0.0106–0.039)	0.0006	T	259	0.2529
<i>AGER</i>	rs2269423	T	-0.014 (-0.025– -0.003)	0.0138	T	332	0.4500
<i>AGER</i>	rs204994	A	0.021 (0.007–0.036)	0.0030	A	272	0.1003
<i>AGER</i>	rs204993	T	0.018 (0.005–0.032)	0.0063	T	303	0.1651
<i>MAP3K5</i>	rs4363056	T	0.013 (0.002–0.0254)	0.0228	T	330	0.2536
<i>KCND2</i>	rs802340	T	-0.019 (-0.036– -0.003)	0.0194	G	221	0.3357

Single nucleotide polymorphisms (SNPs) with p-values ≤ 0.05 in the COPD case-control study are listed. Bold SNPs represent significant replication in both populations. ICGN: International COPD Genetics Network; β : regression coefficient. #: FEV₁/vital capacity (VC) was used for the ICGN family study, with VC determined as the higher of FVC and synchronised VC.

genes were associated with COPD-related phenotypes in the two data sets. To our knowledge, this is the first study to replicate the same SNPs of the *STAT1* and *NFKB1B/SIRT2* genes with COPD-related phenotypes in two populations. These genes have multiple potential relationships with established disease mechanisms in COPD.

The mechanisms by which *STAT1* exhibits its functions are unclear. Progression of COPD is associated with increased numbers of CD8⁺ T-cells and B-cells in the airways, supporting a role for these cells in the pathophysiology of COPD [19]. It is known that CD8⁺ T-lymphocytes express the chemokines CXCL10 and interferon (IFN)- γ [19]. The chemokines act as ligands at the CXCR3 receptor, with expression peaking 8–12 h after stimulation with IFN- γ [20]. This pathway is thought to be dependent on *STAT1* [21]. It is worth noting that TUDHOPE *et al.* [21] observed that this pathway was dexamethasone-resistant, a finding that is consistent with the limited impact of corticosteroids on the inflammatory profile in COPD [22].

The two genes *NFKB1B* and *SIRT2* are closely located (according to HAPMAP). The SNP rs2241704 is located in the flanking regions of the genes and we cannot distinguish them based on the current data. Hence, the observed association to COPD may work through either gene.

Increasing evidence supports a key role for the transcription factor nuclear factor (NF)- κ B in the host response to *Streptococcus pneumoniae* infection [23]. Control of NF- κ B activity is achieved through interactions with the I- κ B family of inhibitors, encoded by the genes *NFKB1A*, *NFKB1B* and *NFKB1E*. COPD patients frequently have lower respiratory tract colonisation with *S. pneumoniae* and this chronic infection may play a role in the pathogenesis of COPD. Functional

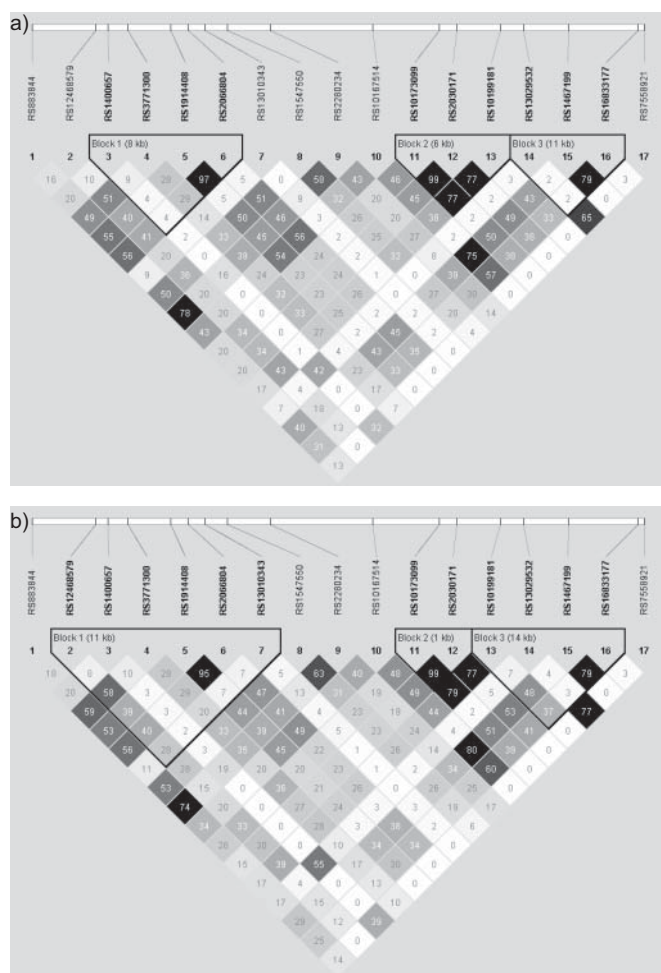


FIGURE 1. Linkage disequilibrium map in *STAT1* gene region in a) the International COPD Genetics Network family-based population and b) case-control population. Values of r^2 ($\times 100$) are shown. ■: $r^2=1$; □: $r^2=0$. Shades of grey represent $0 < r^2 < 1$ (the intensity of the grey is proportional to r^2). Haplotype block structure was estimated with the Haploview program. COPD: chronic obstructive pulmonary disease.

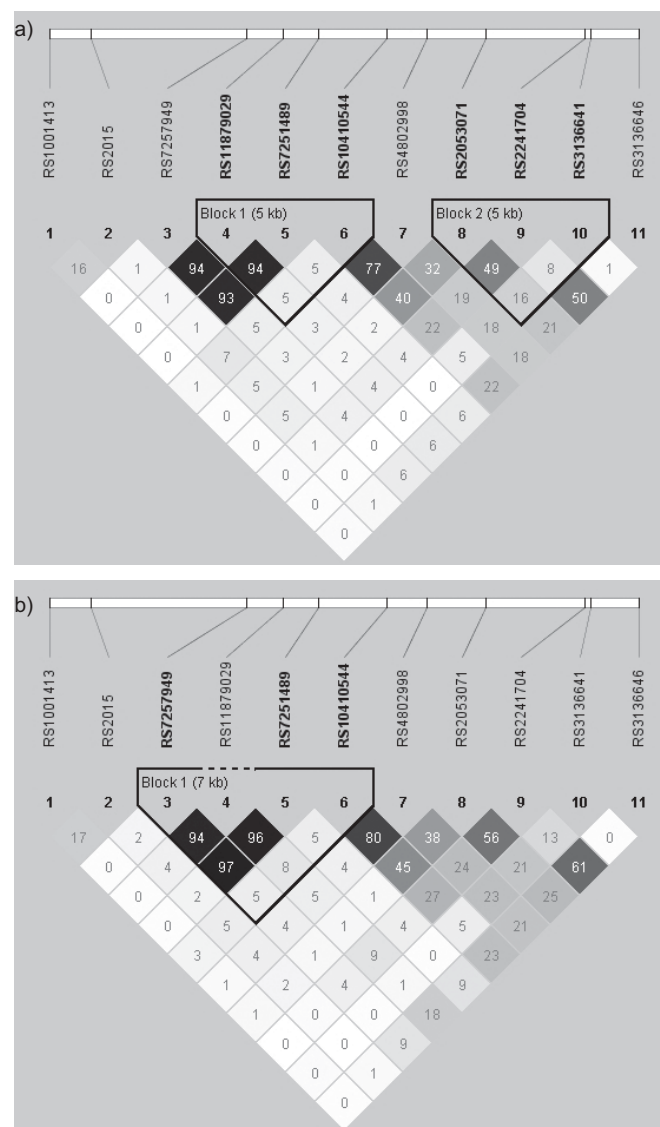


FIGURE 2. Linkage disequilibrium map in *NFKB1B/SIRT2* gene region in a) the International COPD Genetics Network family-based population and b) case-control population. Values of r^2 ($\times 100$) are shown. ■: $r^2=1$; □: $r^2=0$. Shades of grey represent $0 < r^2 < 1$ (the intensity of the grey is proportional to r^2). Haplotype block structure was estimated with the Haploview program. COPD: chronic obstructive pulmonary disease.

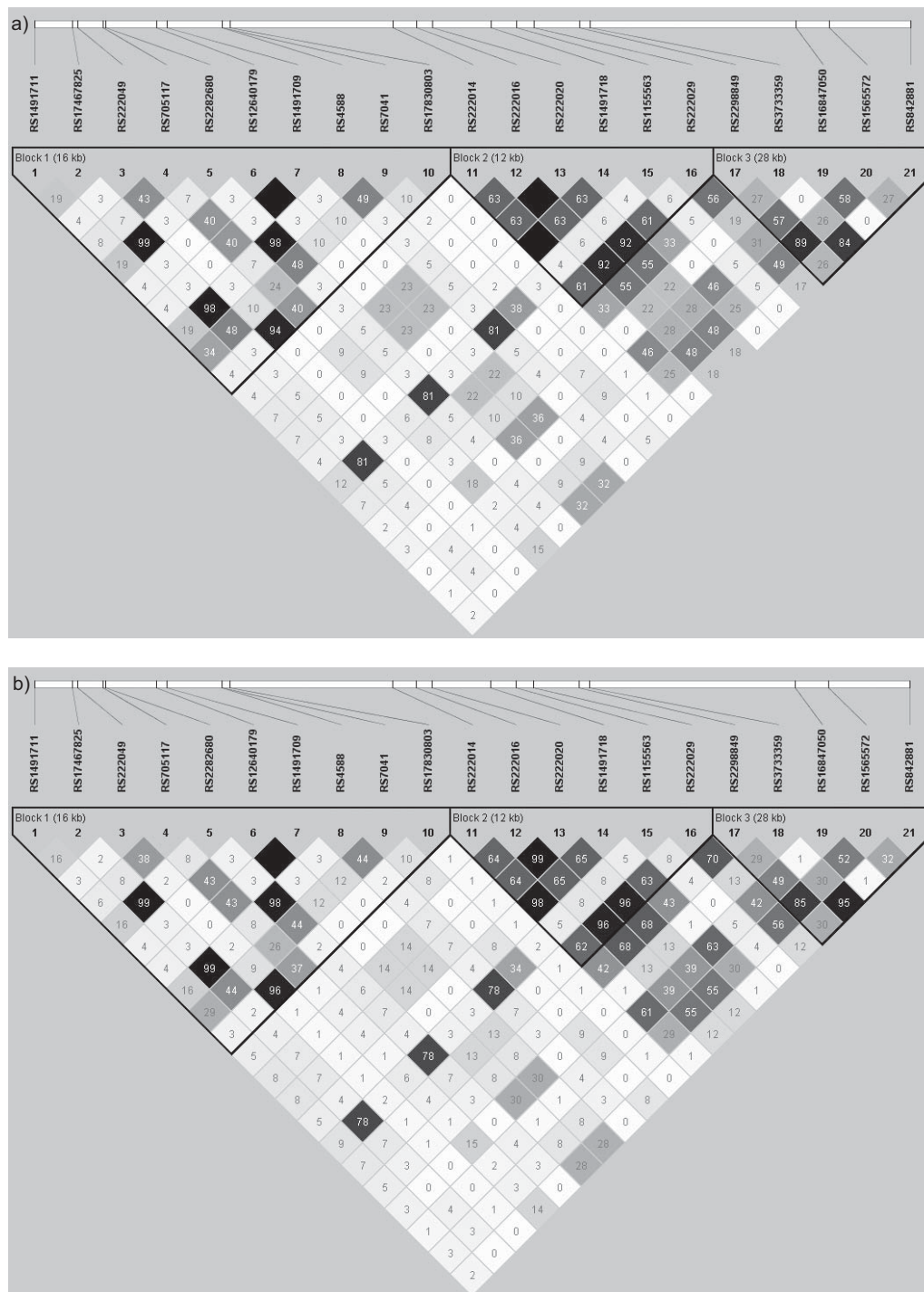


FIGURE 3. Linkage disequilibrium map in GC gene region in a) the International COPD Genetics Network family-based population and b) case-control population. Values of r^2 ($\times 100$) are shown. ■: $r^2=1$; □: $r^2=0$. Shades of grey represent $0 < r^2 < 1$ (the intensity of the grey is proportional to r^2). Haplotype block structure was estimated with the Haploview program. COPD: chronic obstructive pulmonary disease.

polymorphisms in the *NFKB1B* gene may affect the host response to the *S. pneumoniae* infection and, thereby, cause chronic airway inflammation.

The *SIRT2* gene may be related to degenerative processes through the deacetylation of the α -tubulin [24]. The *SIRT* genes are believed to be involved in the ageing process [25, 26].

Emphysema may be regarded as a premature ageing of the lung with loss of elastic fibres [27].

GC encodes the vitamin D binding protein (DBP), which may have several functions. The major function is binding, solubilisation and transport of vitamin D and its metabolites [28]. It is also reported to augment the chemotactic effect of

complement-derived molecules on neutrophils [29]. Neutrophils play an important role in parenchymal destruction and airway inflammation in COPD. Another important function of DBP is its deglycosylation to DBP macrophage-activating factor. The absence of a glycosylated residue at position 420 in *GC2* inhibits this conversion. This may be a partial explanation of its protective effect [30]. Unlike the *GC2* allele, the homozygous *GC1F* phenotype is a significant risk factor for the development of COPD [31, 32]. Although the *GC1F* allele has no effect on the age of onset of COPD, the annual decline in FEV₁ has been reported to be significantly higher in patients with this allele. High-resolution computed tomography showed that *GC1F* allele carriers suffer from more severe emphysema [29]. BLACK and SCRAGG [33] recently analysed serum 25-hydroxyvitamin D₃ concentration, FEV₁ and FVC of 14,091 subjects (aged ≥ 20 yrs) and demonstrated a significant correlation between these parameters.

Several genes were only significantly associated with the COPD-related traits in one of the two studies. This does not necessarily imply that there is no true relationship between these genes and COPD. It especially relates to the genes of which the SNPs reached borderline significance in the other data set, for instance rs2282680 and rs4588, or the *GC* gene (table 3). This makes it less likely that these associations are false-positive. Although the three phenotypes of COPD used in the analyses were identical across the samples (except for VC versus FVC), they are still based solely on spirometry. It is well acknowledged that FEV₁ and FVC far from reflect the whole picture of COPD [34]. Hence, the COPD cases from the two populations might differ with respect to other important characteristics of the disease, such as degree of emphysema and chronic bronchitis, systemic inflammation, body mass index, respiratory failure and rate of exacerbations. More specific COPD-related phenotypes than those applied in the present study might have revealed stronger associations to the genes and enhanced the reproducibility of the relationships across the samples.

In both samples, the participants were examined only once. Hence, we have no data regarding disease progression or rate of lung function decline. Such data might have further strengthened the associations and reproducibility of the findings.

The major exposure in COPD, active smoking, was taken into account in the analyses, based on questionnaire responses. However, distribution of other environmental risk factors, such as occupational airborne exposure, as well as indoor and outdoor air pollution, may differ between the samples and affect the observed associations between the genes and COPD. Passive smoking *in utero* or in early life may cause airway disease in adults [35]. The degree of passive smoking and at what time in life it occurs may vary across the samples and potentially influence any gene–environment interaction effect on COPD.

A common cause of lack of reproducible findings is small sample sizes [36]. In our study, limited statistical power is not likely to explain lack of replication. The size of the population in these studies is appropriate to identify even modest effects.

Random genotyping errors will cause a nondifferential misclassification and bias any association towards nil [37, 38]. However, systematic errors may cause a false positive association. Deviation from HWE in the control group may be a sign of genotyping error [36]. However, in the current study, all the SNPs tested were in HWE in both data sets.

Obviously, the possibility exists that there are true differences in genetic determinants for COPD in various populations. Several genetic COPD association studies have seen inconsistent results between Caucasian and Asian populations [39]. However, it is worth noting that in both the ICGN and the Bergen data sets of the current study, all of the participants were self-reported Caucasians.

The potential weaknesses of the study should be acknowledged. First, the case–control population is from a single centre in Norway, and the age and pack-yrs of smoking were different between the cases and controls. Therefore, we used a logistic regression model for the association analysis, at least partly correcting for age and pack-yrs of smoking. The fact that the case–control population is from a single centre can also be considered as an advantage, because the population is more homogeneous and the possibility of false positive results though population stratification is minimal. The analysis of population stratification using a set of 257 unlinked SNPs showed no evidence for population stratification. Secondly, no SNPs were significantly associated with COPD after correction for multiple comparisons. However, the replication in two independent populations and validation by haplotype analysis suggest that the association results are valid. Although various procedures have been studied for correction of multiple testing, including Bonferroni correction and permutation testing, there has not been an ideal statistical framework to deal with raw p-values from SNP association analyses [40], especially for replication of a previously reported association result. [41]. Here, we used the replication of the results in an independent cohort to validate our primary findings.

In conclusion, we conducted a robust genetic association study and found that variants in the *STAT1*, *NFKB1B/SIRT2* and *GC* genes are likely to contribute to the susceptibility to COPD. Functional tests need to be performed to find the molecular mechanism that drives the genetic association between COPD phenotypes and these three genes.

STATEMENT OF INTEREST

Statements of interest for P.S. Bakke, G. Zhu, X. Kong, A.G.N. Agusti, P.M.A. Calverley, R.D. Levy, S.I. Rennard, W.H. Anderson, D.A. Lomas and E.K. Silverman can be found at www.erj.ersjournals.com/site/misc/statements.xhtml

ACKNOWLEDGEMENTS

The author's affiliations are: P.S. Bakke, Dept of Thoracic Medicine, Haukeland University Hospital and Institute of Medicine, University of Bergen, Bergen, Norway; G. Zhu, GlaxoSmithKline R&D, Research Triangle Park, NC, USA; A. Gulsvik, Dept of Thoracic Medicine, Haukeland University Hospital and Institute of Medicine, University of Bergen; X. Kong, GlaxoSmithKline R&D, Research Triangle Park, NC, USA; A.G.N. Agusti, Hospital Universitari Son Dureta, Fundació Caubet-Cimera and Ciber Enfermedades Respiratorias, Palma, Spain; P.M.A. Calverley, University of Liverpool, Liverpool, UK;

C.F. Donner, Division of Pulmonary Disease, S. Maugeri Foundation, Veruno, Italy; R.D. Levy, Division of Respiratory Medicine and James Hogg iCAPTURE Centre for Cardiovascular and Pulmonary Research, University of British Columbia, St. Paul's Hospital, Vancouver, BC, Canada; B.J. Make and P.D. Paré, both National Jewish Medical and Research Centre, Denver, CO, USA; S.I. Rennard, University of Nebraska, Omaha, NE, USA; J. Vestbo, Dept of Cardiology and Respiratory Medicine, Hvidovre Hospital, Copenhagen, Denmark and University of Manchester, Manchester, UK; E.F.M. Wouters, Dept of Respiratory Medicine, University Hospital Maastricht, the Netherlands; W. Anderson, GlaxoSmithKline R&D, Research Triangle Park, NC, USA; D.A. Lomas, Dept of Medicine, University of Cambridge, Cambridge, UK; E.K. Silverman, The Channing Laboratory and Pulmonary and Critical Care Division, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; and S.G. Pillai, GlaxoSmithKline R&D, Research Triangle Park, NC, USA.

REFERENCES

- Buist AS, McBurnie MA, Vollmer WM, *et al.* International variation in the prevalence of COPD (the BOLD Study): a population-based prevalence study. *Lancet* 2007; 370: 741–750.
- Lopez AD, Mathers CD, Ezzati M, *et al.* Global Burden of Disease and Risk Factors. Washington, The World Bank, 2006.
- Hersh CP, Demeo DL, Lazarus R, *et al.* Genetic association analysis of functional impairment in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2006; 173: 977–984.
- Seifart C, Plagens A. Genetics of chronic obstructive pulmonary disease. *Int J Chron Obstruct Pulmon Dis* 2007; 2: 541–550.
- Ganrot PO, Laurell CB, Erikson S. Obstructive lung disease and trypsin inhibitors in alpha-1-antitrypsin deficiency. *Scand J Clin Lab Invest* 1967; 19: 205–208.
- Hersh CP, Demeo DL, Lange C, *et al.* Attempted replication of reported chronic obstructive pulmonary disease candidate gene associations. *Am J Respir Cell Mol Biol* 2005; 33: 71–78.
- Hall IP, Blakey JD. Genetic association studies in *Thorax*. A guide to assessing the validity of genetic association studies in respiratory disease. *Thorax* 2005; 60: 357–359.
- Zhu G, Warren L, Aponte J, *et al.* The SERPINE2 gene is associated with chronic obstructive pulmonary disease in two populations. *Am J Respir Crit Care Med* 2007; 176: 167–173.
- Johannessen A, Omenaas ER, Eide GE, *et al.* Feasible and simple exclusion criteria for pulmonary reference populations. *Thorax* 2007; 62: 792–798.
- Patel BD, Coxson HO, Pillai SG, *et al.* Airway wall thickening and emphysema show independent familial aggregation in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2008; 178: 500–505.
- International HapMap Project. www.hapmap.org Date last accessed: January 31, 2008. Date last updated: August 18, 2010.
- Pillai SG, Ge D, Zhu G, *et al.* A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. *PLoS Genet* 2009; 5: e1000421.
- Lange C, DeMeo D, Silverman EK, *et al.* PBAT: tools for family-based association studies. *Am J Hum Genet* 2004; 74: 367–369.
- O'Connell JR, Weeks DE. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* 1998; 63: 259–266.
- Horvath S, Xu X, Lake SL, *et al.* Family-based tests for associating haplotypes with general phenotype data: application to asthma genetics. *Genet Epidemiol* 2004; 26: 61–69.
- Schaid DJ, Rowland CM, Tines DE, *et al.* Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am J Hum Genet* 2002; 70: 425–434.
- Barrett JC, Fry B, Maller J, *et al.* Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; 21: 263–265.
- Gabriel SB, Schaffner SF, Nguyen H, *et al.* The structure of haplotype blocks in the human genome. *Science* 2002; 296: 2225–2229.
- Hogg JC, Chu F, Utokaparch S, *et al.* The nature of small-airway obstruction in chronic obstructive pulmonary disease. *N Engl J Med* 2004; 350: 2645–2653.
- Saetta M, Mariani M, Panina-Bordignon P, *et al.* Increased expression of the chemokine receptor CXCR3 and its ligand CXCL10 in peripheral airways of smokers with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2002; 165: 1404–1409.
- Tudhope SJ, Cartley MC, Fenwick PS, *et al.* The role of I κ B kinase 2, but not activation of NF- κ B, in the release of CXCR3 ligands from IFN- γ -stimulated human bronchial epithelial cells. *J Immunol* 2007; 179: 6237–6245.
- Hattotuwa KL, Gizycki MJ, Ansari TW, *et al.* The effects of inhaled fluticasone on airway inflammation in chronic obstructive pulmonary disease: a double-blind, placebo-controlled biopsy study. *Am J Respir Crit Care Med* 2002; 165: 1592–1596.
- Chapman SJ, Khor CC, Vannberg FO, *et al.* I κ B genetic polymorphisms and invasive pneumococcal disease. *Am J Respir Crit Care Med* 2007; 176: 181–187.
- Rajendrasozhan S, Yang SR, Kinnula VL, *et al.* SIRT1, an anti-inflammatory and antiaging protein, is decreased in lungs of patients with chronic obstructive pulmonary disease. *Am J Respir Dis* 2008; 177: 861–870.
- Outeiro TF, Kontopoulos E, Altmann SM, *et al.* Sirtuin 2 inhibitors rescue α -synuclein-mediated toxicity in models of Parkinson's disease. *Science* 2007; 317: 516–519.
- North BJ, Marshall BL, Borra MT, *et al.* The human *Sir2* ortholog, *SIRT2*, is an NAD⁺-dependent tubulin deacetylase. *Mol Cell* 2003; 11: 437–444.
- Milne JC, Denu JM. The Sirtuin family: therapeutic targets to treat diseases of aging. *Curr Opin Chem Biol* 2008; 12: 11–17.
- Daiger SP, Schanfield MS, Cavalli-Sforza LL. Group-specific component (Gc) proteins bind vitamin D and 25-hydroxyvitamin D. *Proc Natl Acad Sci USA* 1975; 72: 2076–2080.
- Speeckaert M, Huang G, Delanghe JR, *et al.* Biological and clinical aspects of the vitamin D binding protein (Gc-globulin) and its polymorphism. *Clin Chim Acta* 2006; 372: 33–42.
- Horne SL, Cockcroft DW, Dosman JA. Possible protective effect against chronic obstructive airways disease by the GC2 allele. *Hum Hered* 1990; 40: 173–176.
- Ito I, Nagai S, Hoshino Y, *et al.* Risk and severity of COPD is associated with the group-specific component of serum globulin 1F allele. *Chest* 2004; 125: 63–70.
- Ishii T, Keicho N, Teramoto S, *et al.* Association of Gc-globulin variation with susceptibility to COPD and diffuse panbronchiolitis. *Eur Respir J* 2001; 18: 753–757.
- Black PN, Scragg R. Relationship between serum 25-hydroxyvitamin D and pulmonary function in the Third National Health and Nutrition Examination Survey. *Chest* 2005; 128: 3792–3798.
- Celli BR. Roger S. Mitchell lecture. Chronic obstructive pulmonary disease phenotypes and clinical relevance. *Proc Am Thorac Soc* 2006; 3: 461–465.
- Skorge TD, Eagan TM, Eide GE, *et al.* The adult incidence of asthma and respiratory symptoms by passive smoking in uterus or in childhood. *Am J Respir Crit Care Med* 2005; 172: 61–65.
- Silverman EK, Palmer LJ. Case-control association studies for the genetics of complex respiratory diseases. *Am J Respir Cell Mol Biol* 2000; 22: 645–648.
- Rothman KJ, Greenland S. Precision and validity in epidemiological studies. *In: Rothman KJ, Greenland S, eds.*

- Modern Epidemiology. Philadelphia, Lippincott-Raven, 1998; pp. 115–134.
- 38** Bakke PS. Designing studies for diagnostic tests. *Clin Respir J* 2008; 2: 72–75.
- 39** Molfino N. Current thinking on genetics of chronic obstructive pulmonary disease. *Curr Opin Pulm Med* 2007; 13: 107–113.
- 40** Roeder K, Bacanu SA, Sonpar V, *et al.* Analysis of single-locus tests to detect gene/disease associations. *Genet Epidemiol* 2005; 28: 207–219.
- 41** Webster PM, Lorimer EG, Man SF, *et al.* Pulmonary function in identical twins: comparison of nonsmokers and smokers. *Am Rev Respir Dis* 1979; 119: 223–228.