ERJ Express. Published on August 6, 2010 as doi: 10.1183/09031936.00026410

Surfactant protein-B polymorphisms, pulmonary function and COPD in 10,231 individuals

Marie Bækvad-Hansen,[1] Børge G Nordestgaard,[1,2] Morten Dahl[1,*]

[1]Department of Clinical Biochemistry, Herlev Hospital, Copenhagen University Hospital; [2]The Copenhagen City Heart Study, Bispebjerg Hospital, Copenhagen University Hospital; [1-2]Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark.

Running title: *SP-B* polymorphisms and COPD

*Correspondence to: Morten Dahl, MD., DMSc., Department of Clinical Biochemistry, Herlev Hospital, Copenhagen University Hospital, Herlev Ringvej 75, DK-2730 Herlev, Denmark. Phone: +45 2737 8010, Fax: +45 4488 3311, e-mail: mordah02@heh.regionh.dk

Word count manuscript: 2685
Word count abstract: 198
Character count title: 85

Keywords: chronic obstructive pulmonary disease; genetics; asthma; interstitial lung disease; pneumonia; lung cancer.

Abstract

The *surfactant protein-B (SP-B)* gene may influence COPD and thus personalised medicine. We tested whether functional polymorphisms in *SP-B* (rs1130866=1580T>C, rs2077079=-18A>C and rs3024791=-384G>A) associate with reduced lung function and risk of COPD in the general population.

We genotyped 10,231 individuals from the adult Danish general population, and recorded spirometry and hospital admissions due to COPD. Because we previously found an association between the rare *SP-B* 121ins2 mutation and COPD among smokers, we stratified the analyses for smoking status.

None of the individual *SP-B* genotypes or genotype combinations were associated with reduced FEV_1%predicted, FVC%predicted and FEV_1/FVC overall, or among smokers separately (p=0.25-0.99). The odds ratio for spirometry defined COPD did not differ from 1.0 for any of the *SP-B* genotypes or genotype combinations overall, or among smokers separately (p=0.17 to 0.78). Similar results were obtained for hospitalisation due to COPD (p=0.07 to 0.93); we could exclude overall hazard ratios for heterozygotes of 1.18 to 1.21 and for homozygotes of 1.25 to 1.57 or larger for all three polymorphisms.

In conclusion, functional polymorphisms in the *SP-B* gene are not associated with reduced lung function or risk of COPD, making it unlikely that these variants will be useful in personalised medicine.

Introduction

Chronic obstructive pulmonary disease (COPD) is one of the leading causes of death worldwide and the number of deaths due to COPD is expected to rise in the future [1]. Smoking is the main risk factor for development of COPD, however far from all smokers develop the disease. This suggests that other factors such as genetic background may play a role in susceptibility to COPD, and thus be useful in personalised medicine. A variety of genes have been linked with risk of COPD; among these the surfactant protein genes [2-4].

Surfactant proteins are essential components of the lung surfactant layer which covers the terminal airways. The lung surfactant consists of phospholipids, cholesterol, and proteins and is primarily produced by alveolar type II cells [5,6]. It forms a thin lipid layer on the surface of alveoli, which reduces surface tension and prevents alveoli from collapse during expiration. Upon exocytosis from alveolar type II cells, surfactant initially exists as multilayered vesicular structures in the epithelial lining fluid. From these structures it spreads to the surface as alveoli are extended and compressed during breathing.

Surfactant protein-B (SP-B) is one of the four known surfactant proteins in humans. It is important for the formation of lamellar bodies and correct assembly of the surfactant layer [7]. Lack of SP-B cause fatal respiratory distress syndrome in newborns [8], and genetic markers in and around the *SP-B* gene have been associated with a spectrum of pulmonary diseases including COPD [2-4,9-13]. Three functional polymorphisms in *SP-B*, rs1130866, rs2077079 and rs3024791, have been associated with risk and/or severity of COPD. Rs1130866 abolishes an N-linked glycosylation site in *SP-B*, while the two promoter polymorphisms, rs2077079 and rs3024791 alter *SP-B* transcription levels [3,14-18].

We hypothesised, that the three common functional polymorphisms (rs1130866, rs2077079 and rs3024791) in the *SP-B* gene are associated with reduced lung function and risk of

COPD in the general population. To test this hypothesis, we genotyped 10,231 individuals from the adult Danish general population, and recorded spirometry and hospital admissions due to COPD. We calculated odds and hazard ratios to assess risk of COPD according to SP-B genotype, and we used power calculation to illustrate the maximal risk of COPD we potentially could have overlooked.

Materials and methods

Subjects

We genotyped 10,231 individuals from the Copenhagen City Heart Study, a prospective general population study of individuals selected based on the Central Population Register Code to reflect the adult Danish population aged 20 to 80+ years [19]. The Copenhagen City Heart Study was initiated in 1976-1978 with follow-up examinations in 1981-1983, 1991-1994 and in 2001-2003. DNA for genotyping was isolated from participants attending the 1991-1994 and/or 2001-2003 examinations. The study was approved by the local ethical committee: Nos. 100.2039/91 and 01-421/94, Copenhagen and Frederiksberg committee. All participants gave written informed consent. All participants were Whites of Danish descent

Pulmonary function testing and COPD diagnoses

FEV_1 and FVC were determined with a dry wedge spirometer (Vitalograph; Maids Moreton, Buckinghamshire, UK). Each spirometry was performed in triplicate and results accepted only if variation between the two best performing of these were less than 5%. The best results were used for calculation of FEV_1 % predicted and FVC % predicted by using multiple regressions with age and height as covariates on never smokers for men and women separately [19,20]. Spirometry defined COPD was $FEV_1/FVC<0.7$ and FEV_1 <80% of predicted [4]. If individuals with asthma were excluded from this definition and/or individuals with GOLD stage I ($FEV_1/FVC<0.7$ and FEV_1 >80% of predicted) were included, the results were similar to those presented. Information on hospitalisation due to COPD (ICD8:491-492, ICD10:J41-J44) was collected in the national Danish Patient Registry, which covers all hospitals in Denmark from 1976 through 2009.

Other pulmonary diseases

Asthma was defined as an affirmative response to the question “do you have asthma?” Chronic bronchitis was defined as an affirmative response to the question “Do you bring up phlegm at least

3 months continuously every year?" Information on pneumonia, interstitial lung disease and lung cancer was collected in the national Danish Patient Registry and the Danish Cancer Registry. Pneumonia (ICD 8:480-486, ICD10:J12-J18), Interstitial lung disease (ICD10:J84) and lung cancer (ICD7:162.0-2, 162.4-7, 163.0, 164.0, 199.2, 462.1-4, 464.4, 962.1-2, 962.4-6, 963.0, 964.0, ICD10: C33-C34, C37-C38, D02.1-2).

Genotyping

Genomic DNA was isolated from frozen whole blood (Qiagen, Hilden, Germany). The genotype analysis was performed in 2007 by use of the ABI PRISM® 7900HT Sequence Detection System (Applied Biosystems Inc., Foster City, CA, USA). Primers and probes for the TaqMan assays are listed in supplementary Table 1. The TaqMan analysis was validated by sequencing of a subsample of the participants.

Statistical analysis

Statistical analyses were performed with STATA/S.E version 10.0 (StataCorp LP, College Station, Texas, USA). A two-sided $P<0.05$ was considered significant. From the three polymorphisms, we generated all possible genotype combinations. We used Kruskal-Wallis test and Pearson's χ^2 test for differences in characteristics between *SP-B* genotypes (Table 1). Main effects of *SP-B* genotypes in predicting FEV_1 % predicted, FVC % predicted and FEV_1/FVC were examined using the test for trends, while associations of *SP-B* genotype combinations with lung function were tested in ANOVA models. Odds ratios for COPD on spirometry by *SP-B* genotype/genotype combination were by logistic regression models including age, sex, packyears, passive smoking, occupational dust/fumes, education, and occupation. Hazard ratios for COPD hospitalisation during up to 33 years of follow-up by *SP-B* genotype/genotype combination were determined by Cox regression models with age as time scale, adjusted for age, sex, packyears passive smoking, occupational dust/fumes, education, and occupation. We used NCSS-PASS to calculate the low and high

odds/hazard ratios, which we have 90% power to exclude at 2-sided probability values <0.05. Linkage disequilibrium between polymorphisms was estimated by Lewontins D' using STATA's pwld function.

Results

Characteristics of the participants are shown in Table 1. There were no differences in sex, age, smoking status or packyears of tobacco smoked for any of the *SP-B* genotypes. The distributions of *SP-B* genotypes for the three polymorphisms were in Hardy-Weinberg equilibrium (rs1130866: p=0.26, rs2077079: p=0.96, rs3024791: p=0.32). Pair wise linkage disequilibrium for the three polymorphisms was determined by D'. Rs2077079 and rs3024791 were in tight linkage disequilibrium (D'=0.97), whereas there was low linkage disequlibrium between rs1130866 and either of rs2077079 or rs3024791 (D'=0.21 and 0.18 respectively). The relationship between spirometry defined COPD and COPD hospitalisation according to gender and age is presented in Table 2. The proportion of men tended to be higher among those with spirometry defined COPD and age greater than 66 years, while the proportion of women tended to be higher among those with COPD hospitalisation and age less than 67. The proportions of women and men were similar among individuals with spirometry defined COPD or COPD hospitalisation.

Lung function

None of the individual *SP-B* genotypes or genotype combinations were associated with reduced FEV_1 % predicted, FVC % predicted and FEV_1/FVC (p=0.34 to 0.94) (Figure 1). When stratifying for smoking status, the p-value for *SP-B* genotype combinations reached significance for FEV_1 % predicted and FVC % predicted among never smokers (Figure 2). On post-hoc analysis the *SP-B* TCAAGG and TCCCGG vs TCACGG genotypes were associated with increased FEV_1 % predicted (p=0.02 and p=0.01, respectively) and FVC % predicted (p=0.02 and p=0.003) among never smokers. Conversely the *SP-B* CCCCGA vs TCACGG genotype was associated with reduced FEV_1 % predicted (p=0.007), FVC % predicted (p=0.04), and FEV_1/FVC (p=0.01) among never smokers. These results are not biologically plausible and they could not be confirmed when testing spirometry defined COPD or COPD hospitalisation. Among ever smokers, none of the individual

SP-B genotypes or genotype combinations were associated with reduced FEV_1 % predicted, FVC % predicted and FEV_1/FVC (p = 0.25 to 0.99) (Supplementary Figure 1).

Risk of COPD

The odds ratio for spirometry defined COPD did not differ from 1.0 for any *SP-B* genotype or genotype combination (Figure 3). The risk for COPD hospitalisation did neither differ from 1.0 for any of the *SP-B* genotypes or genotype combinations (Figure 3). We had 90% statistical power to exclude odds ratios for spirometry defined COPD for heterozygotes of 1.22 to 1.24 and for rare homozygotes of 1.30 to1.96 or larger for all three polymorphisms (Figure 3). Likewise, we had 90% statistical power to exclude hazard ratios for COPD hospitalisation for heterozygotes of 1.18 to 1.21 and for rare homozygotes of 1.25 to 1.57 or larger for all three polymorphisms.

When stratifying for smoking status, the p-value for the rs1130866 polymorphism reached significance for COPD hospitalisation among never smokers (Figure 4); however, none of the individual rs1130866 CC and TC genotypes differed significantly from 1.0, and the result could not be confirmed when analysing spirometry-defined COPD. The odds ratio for spirometry defined COPD was increased in *SP-B* TTACGG and TCAAGA vs TCACGG among never smokers (odds ratios and 95%CI: 2.5 (1.2-5.5) and 2.7 (1.1-6.7), respectively), and for COPD hospitalisation in *SP-B* CCAAAA vs TCACGG among never smokers (hazard ratio 8.6 (1.7-42.6)). None of these results could be confirmed when analysing the other COPD outcome of the study or when analysing the lung function. We therefore interpret these findings as likely spurious results.

Because the rs1130866 and the rs2077079 polymorphisms have been associated with severity of COPD and COPD exacerbations [3,16], we examined the prevalence of COPD GOLD stages by *SP-B* genotype. The prevalence of COPD GOLD stage II and COPD GOLD stage III/IV did not differ by any of the *SP-B* genotypes (p = 0.16 to 0.86).

Risk of other lung diseases

We also tested whether the three *SP-B* polymorphisms were associated with asthma, interstitial lung disease, pneumonia, chronic bronchitis or lung cancer (Supplementary Table 2). None of the *SP-B* genotypes associated with any of the abovementioned lung diseases; except that the rs3024791 AA vs. GG genotype did show an association with 1.8-fold increased risk of asthma.

Discussion

Genetic variation in *SP-B* has been linked to COPD in several association studies [3,9,16,18] and thus could have importance for the development of COPD in the general population. We determined whether three functional polymorphisms in *SP-B* were associated with poor lung function and COPD in a large homogenous Danish population sample. We found with significant power that these polymorphisms were not associated with lung function or risk of COPD overall, or among smokers. This makes it unlikely that the genetic variants can be used clinically to assess risk of COPD or to identify COPD subgroups for tailored therapy.

The C allele of the rs113086 polymorphism has previously, in studies with <1,000 participants, been associated with increased risk of COPD, severity of airway obstruction and disease exacerbation in COPD patients [3,14,16,18]. In contrast, our population based study with > 10,000 participants indicate that the rs1130866 polymorphism does not affect risk of COPD, severity of COPD, or risk of any other additional pulmonary disorder overall or among smokers. We did find a trend towards lower risk of COPD hospitalisation among never smokers. However, this result could not be confirmed when analysing spirometry defined COPD or lung function. We therefore interpret this as a likely spurious finding. Our data on lung function are in accordance with a previous study on healthy males, which showed no association between rs1130866 and FEV_1 % predicted or FVC % predicted [21].

The *SP-B* promoter variations rs2077079 and rs3024791 have previously, in studies with <400 participants, been associated with severity of airway obstruction and exacerbations in COPD patients, but not to risk of developing COPD [3,16]. Both polymorphisms alter transcription of the *SP-B* gene and may be associated with altered levels of SP-B in the airways. We did not find any overall association of the two promoter variants with lung function, risk of COPD, or COPD hospitalisation. We have previously shown, that partial SP-B deficiency due to the rare 121ins2

mutation is associated to reduced pulmonary function and increased risk of COPD among smokers [4]. However, when stratifying our data for smoking status, none of the two promoter polymorphisms were associated with reduced pulmonary function or risk of COPD among smokers.

It has been suggested that the rs1130866 polymorphism may require other interacting factors to alter the pulmonary phenotype. Guo et al suggested that gene-gene and gene-environment interaction may be of importance [3] and Hersh et al. found rs113086 to be associated with COPD only in a statistical model with the presence of a gene-smoking interaction term [18]. To mimic gene-gene interactions and gene-environment interactions, we assessed lung function and calculated COPD risk estimates for combinations of rs1130866, rs2077079 and rs3024791 stratified by smoking status. When stratifying our analyses by smoking status, we were unable to show association of *SP-B* genotype combinations with lung function and risk of COPD among smokers. Among never smokers we did find association between certain genotype combinations and lung function or COPD. However, these results may not be biologically plausible and they could not be confirmed using lung function or the other COPD endpoint of the study. If correction for 60 and 30 multiple comparisons was performed for the analyses of *SP-B* combinations and lung function (Figure 2) and *SP-B* combinations and COPD (Figure 4) respectively, none of the results observed among never smokers would be of statistical significance. We therefore interpret these findings as likely spurious results. The lack of overall association of *SP-B* polymorphisms with pulmonary function and disease in our study is supported by recent genome-wide association studies and by a novel Dutch population/case-control study [22-24].

As variations in *SP-B* have been associated to other respiratory diseases than COPD [9,10,25], we also tested for association between the three polymorphisms and risk of common pulmonary diseases such as asthma, pneumonia and lung cancer. We found no consistent association between *SP-B* genotypes and any of the lung diseases we examined except for asthma.

As did previous reports on surfactant protein-C [26,27], a family relative to surfactant protein-B, we did find an association between *SP-B* rs3024791 rare vs. common homozygosity and asthma. We could not confirm this result when analysing other *SP-B* genotypes, and further studies will be needed to conclusively determine whether the *SP-B* rs3024791 AA genotype is associated with increased asthma risk.

Some misclassification of spirometry defined COPD is possible, and this could limit the subgroup analysis according to GOLD classification. The pulmonary function tests used to define COPD were not performed post bronchodilator due to the large number of subjects included in the study and our limited funds. However, excluding individuals with asthma from this definition did not substantially alter our results. Furthermore, in the 1991-1994 survey those individuals who had FEV1/FVC less than 0.7 had post bronchodilator spirometry performed. If analyses on pulmonary function and COPD were confined to this subgroup, we found no difference in FEV1 % predicted or FEV1/FVC, or in COPD prevalences according to SP-B genotype or genotype combination. Lack of association between any of the SP-B polymorphisms with pulmonary function or disease as opposed to previous findings could be due to different genotype frequencies among the studied populations. Hersh et al reported a carrier frequency for the rs1130866 T allele of 0.44 for controls and 0.46 for cases [18]. We found the overall frequency of the T allele to be 0.54. Foreman et al reported carrier frequencies for rs3024791 GG/AG/AA genotypes of 0.73/0.25/0.02 [16]. Our results were very similar (0.75/0.23/0.02). Gou et al reported a carrier frequency for the rs1130866 C allele of 0.67 for controls and 0.82 for cases [3]. We found an overall carrier frequency for the rs1130866 C allele of 0.71. Bias caused by investigator knowledge of disease or risk factor status seems unlikely, because our sample was selected from the general population and because genotyping of our sample was performed without investigator knowledge of disease status or lung function test results.

In conclusion, we find with significant power that common functional polymorphisms in the *SP-B* gene are not associated with reduced lung function or risk of COPD in the Danish general population overall or among smokers. This makes it unlikely that these genetic variants will have a role in personalised medicine. Though our results are based on individuals of Danish/European descent, these polymorphisms are prevalent in many populations and our results may apply to other parts of the World.

References

1 Murray CJ , Lopez AD.Alternative projections of mortality and disability by cause 1990-2020: Global Burden of Disease Study. *Lancet* 1997;349:1498-1504.

2 Hersh CP, DeMeo DL, Lazarus R, Celedon JC, Raby BA, Benditt JO, Criner G, Make B, Martinez FJ, Scanlon PD, Sciurba FC, Utz JP, Reilly JJ, Silverman EK.Genetic association analysis of functional impairment in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2006;173:977-984.

3 Guo X, Lin HM, Lin Z, Montano M, Sansores R, Wang G, DiAngelo S, Pardo A, Selman M, Floros J.Surfactant protein gene A, B, and D marker alleles in chronic obstructive pulmonary disease of a Mexican population. *Eur Respir J* 2001;18:482-490.

4 Baekvad-Hansen M, Dahl M, Tybjaerg-Hansen A, Nordestgaard BG.Surfactant protein-B 121ins2 heterozygosity, reduced pulmonary function, and chronic obstructive pulmonary disease in smokers. *Am J Respir Crit Care Med* 2010;181:17-20.

5 Andreeva AV, Kutuzov MA, Voyno-Yasenetskaya TA.Regulation of surfactant secretion in alveolar type II cells. *Am J Physiol Lung Cell Mol Physiol* 2007;293:L259-L271.

6 Dietl P , Haller T.Exocytosis of lung surfactant: from the secretory vesicle to the air-liquid interface. *Annu Rev Physiol* 2005;67:595-621.

7 Hawgood S.Surfactant protein B: structure and function. *Biol Neonate* 2004;85:285-289.

8 Wilder MA.Surfactant protein B deficiency in infants with respiratory failure. *J Perinat Neonatal Nurs* 2004;18:61-67.

9 Floros J, Lin HM, Garcia A, Salazar MA, Guo X, DiAngelo S, Montano M, Luo J, Pardo A, Selman M.Surfactant protein genetic marker alleles identify a subgroup of tuberculosis in a Mexican population. *J Infect Dis* 2000;182:1473-1478.

10 Lin Z, Pearson C, Chinchilli V, Pietschmann SM, Luo J, Pison U, Floros J.Polymorphisms of human SP-A, SP-B, and SP-D genes: association of SP-B Thr131Ile with ARDS. *Clin Genet* 2000;58:181-191.

11 Quasney MW, Waterer GW, Dahmer MK, Kron GK, Zhang Q, Kessler LA, Wunderink RG.Association between surfactant protein B + 1580 polymorphism and the risk of respiratory failure in adults with community-acquired pneumonia. *Crit Care Med* 2004;32:1115-1119.

12 Floros J, Veletza SV, Kotikalapudi P, Krizkova L, Karinch AM, Friedman C, Buchter S, Marks K.Dinucleotide repeats in the human surfactant protein-B gene and respiratory-distress syndrome. *Biochem J* 1995;305 (Pt 2):583-590.

13 Seifart C, Plagens A, Brodje D, Muller B, von WP, Floros J.Surfactant protein B intron 4 variation in German patients with COPD and acute respiratory failure. *Dis Markers* 2002;18:129-136.

14 Hu R, Xu Y, Zhang Z.Surfactant protein B 1580 polymorphism is associated with susceptibility to chronic obstructive pulmonary disease in Chinese Han population. *J Huazhong Univ Sci Technolog Med Sci* 2004;24:216-8, 238.

15 Steagall WK, Lin JP, Moss J.The C/A(-18) polymorphism in the surfactant protein B gene influences transcription and protein levels of surfactant protein B. *Am J Physiol Lung Cell Mol Physiol* 2007;292:L448-L453.

16 Foreman MG, DeMeo DL, Hersh CP, Carey VJ, Fan VS, Reilly JJ, Shapiro SD, Silverman EK.Polymorphic variation in surfactant protein B is associated with COPD exacerbations. *Eur Respir J* 2008;32:938-944.

17 Thomas KH, Meyn P, Suttorp N.Single nucleotide polymorphism in 5'-flanking region reduces transcription of surfactant protein B gene in H441 cells. *Am J Physiol Lung Cell Mol Physiol* 2006;291:L386-L390.

18 Hersh CP, DeMeo DL, Lange C, Litonjua AA, Reilly JJ, Kwiatkowski D, Laird N, Sylvia JS, Sparrow D, Speizer FE, Weiss ST, Silverman EK.Attempted replication of reported chronic obstructive pulmonary disease candidate gene associations. *Am J Respir Cell Mol Biol* 2005;33:71-78.

19 Dahl M, Bowler RP, Juul K, Crapo JD, Levy S, Nordestgaard BG.Superoxide dismutase 3 polymorphism associated with reduced lung function in two large populations. *Am J Respir Crit Care Med* 2008;178:906-912.

20 Lee J, Nordestgaard BG, Dahl M. Elevated ACE activity is not associated with asthma, COPD, and COPD co-morbidity. *Respir Med* 2009;103:1286-1292.

21 Raleigh SM, Davies BM, Cleal D, Ribbans WJ.No association between coding polymorphism within Exon 4 of the human surfactant protein B gene and pulmonary function in healthy men. *J Physiol Sci* 2007;57:199-202.

22 Wilk JB, Chen TH, Gottlieb DJ, Walter RE, Nagle MW, Brandler BJ, Myers RH, Borecki IB, Silverman EK, Weiss ST, O'Connor GT.A genome-wide association study of pulmonary function measures in the Framingham Heart Study. *PLoS Genet* 2009;5:e1000429-

23 Pillai SG, Ge D, Zhu G, Kong X, Shianna KV, Need AC, Feng S, Hersh CP, Bakke P, Gulsvik A, Ruppert A, Lodrup Carlsen KC, Roses A, Anderson W, Rennard SI, Lomas DA, Silverman EK, Goldstein DB.A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. *PLoS Genet* 2009;5:e1000421-

24 van Diemen CC, Postma DS, Aulchenko YS, Snijders PJ, Oostra BA, van Duijn CM, Boezen HM.Novel strategy to identify genetic risk factors for COPD severity: a genetic isolate. *Eur Respir J* 2010;35:768-775.

25 Seifart C, Seifart U, Plagens A, Wolf M, von WP.Surfactant protein B gene variations enhance susceptibility to squamous cell carcinoma of the lung in German patients. *Br J Cancer* 2002;87:212-217.

26 Erpenbeck VJ, Schmidt R, Gunther A, Krug N, Hohlfeld JM. Surfactant protein levels in bronchoalveolar challenge in patients with asthma. *Allergy* 2006;61:598-604.

27 Baekvad-Hansen M, Nordestgaard BG, Tybjaerg-Hansen A, Dahl M.Two novel mutations in surfactant protein-C, lung function and obstructive lung disease. *Respir Med* 2010;104:418-425

Legends to figures

Figure 1 - Lung function according to *SP-B* genotypes and genotype combinations. Values are means and standard errors. P-values for *SP-B* genotypes and genotype combinations were by test for trends and ANOVA, respectively. Numbers are slightly less than all individuals genotyped as not all participants had spirometry performed.

Figure 2 - Lung function according to *SP-B* genotypes and genotype combinations among never smokers. Values are means and standard errors. P-values for *SP-B* genotypes and genotype combinations were by test for trends and ANOVA, respectively. Numbers are slightly less than all individuals genotyped as not all participants had spirometry performed.

Figure 3 - Risk of COPD according to *SP-B* genotypes and genotype combinations. COPD spirometry = $FEV_1/FVC<0.7$ and $FEV_1<80\%$ of predicted. COPD hospitalisation = ICD8:491-492 or ICD10:J41-J44. Odds ratios for spirometry defined COPD are by logistic regression. Hazard ratios for COPD hospitalisation are by Cox regression. Multivariate adjusted models allowed for age, sex, packyears, passive smoking, occupational dust/fumes, education, and occupation. P-values for *SP-B* genotypes and genotype combinations were by test for trends and ANOVA, respectively. The 90% power indicates the odds ratios that can be detected in this study at two-sided $p<0.05$. Numbers for the analysis of spirometry defined COPD are slightly less than for COPD hospitalisation, because not all individuals had spirometry performed.

Figure 4- Risk of COPD according to *SP-B* genotypes and genotype combinations, stratified by smoking status. Ever smoker was current and former smokers. COPD spirometry = $FEV_1/FVC <0.7$ and $FEV_1<80\%$ of predicted. COPD hospitalisation = ICD8:491-492 or ICD10:J41-J44. Odds ratios

for spirometry defined COPD are by logistic regression. Hazard ratios for COPD hospitalisation are by Cox regression. The models allowed for age, sex, packyears, passive smoking, occupational dust/fumes, education, and occupation. P-values for *SP-B* genotypes and genotype combinations were by test for trends and ANOVA, respectively. Numbers for the analysis of spirometry defined COPD are slightly less than for COPD hospitalisation, because not all individuals had spirometry performed.

Table 1. Characteristics of participants by *SP-B* genotype

Characteristics	All	rs1130866, 1580T>C			p value	rs2077079, -18A>C			p value	rs3024791, -384G>A			p value
		TT	TC	CC		AA	AC	CC		GG	GA	AA	
Number (%)	10231	2963 (29)	5138 (50)	2130 (21)		3815 (37)	4860 (48)	1556 (15)		7724 (75)	2319 (23)	188 (2)	
Women (%)	56	45	45	43	0.49	45	44	44	0.86	44	44	48	0.50
Age (years)	59(45-69)	58 (44-69)	59 (45-69)	58 (45-69)	0.49	59 (45-69)	58 (45-69)	59 (45-69)	0.76	58 (45-69)	59 (45-69)	59 (41-68)	0.38
Smoking (%)													
Never	24	25	23	24	0.15	25	24	23	0.18	24	24	24	0.97
Ever	76	75	77	76		75	76	77		76	76	76	
Packyears	24(10-40)	23 (10-39)	24 (10-40)	24 (10-40)	0.48	24 (10-40)	24 (10-40)	23 (10-40)	0.49	24 (10-40)	23 (10-38)	25 (11-38)	0.59
Passive smoking (%)	79	79	79	79	0.71	79	79	80	0.86	79	79	82	0.83
Occupational exposure to dust/fumes (%)	17	17	17	17	0.84	18	17	17	0.12	18	17	15	0.46
Education					0.86				0.39				0.09
None (%)	23	23	23	23		23	24	21		24	21	24	
<3 years (%)	52	52	52	51		52	52	53		52	52	50	
>3 years (%)	25	24	25	25		25	25	25		24	27	26	
Occupation					0.97				0.46				0.23
No paid job (%)	9	9	9	9		9	10	8		9	9	12	
Employed (%)	80	81	81	80		81	80	82		80	82	81	
Self-employed (%)	10	9	10	10		10	10	9		10	9	7	
Asthma (%)	6	6	6	7	0.39	6	6	5	0.35	6	6	10	0.08
GOLD II (%)	9	9	9	9	0.53	9	10	9	0.26	9	9	6	0.23
GOLD III/IV (%)	4	5	4	4	0.21	4	4	4	0.86	4	4	7	0.16

Values are median (interquartile range) for continuous variables and frequencies for categorical variables. The p-value was calculated by Kruskal-Wallis test for continuous variables and Pearson's χ^2 test for categorical variables. Packyears of tobacco smoked were based on ever smokers. Passive smoking = affirmative response to the questions "does anyone in your household smoke?" or "how many hours per day are you subjected to passive smoke?" Occupational exposure to dust or smoke = affirmative response to the question "Have you been exposed to occupational dust or fumes?" Asthma = affirmative response to the question "Do you have asthma?"

Table 2. Patients with COPD according to gender and age.

		Women		Men	
	All	67< years	67≥ years	66< years	66≥ years
COPD spirometry, No (%)	1327 (100)	296(22)	309 (23)	340 (26)	382 (29)
COPD hospitalisation, No (%)	1186 (100)	347(29)	297 (25)	278 (23)	264 (22)
COPD any, No (%)	1936 (100)	501(26)	462 (24)	492 (25)	481 (25)

Values are number of individuals (%). COPD spirometry = $FEV_1/FVC<0.7$ and $FEV_1<80\%$ of predicted. COPD hospitalisation = ICD8:491-492 or ICD10:J41-J44. Median ages used for stratification were determined among individuals who had spirometry defined COPD or COPD hospitalisation (=COPD any) for women and men separately.

Figure 1

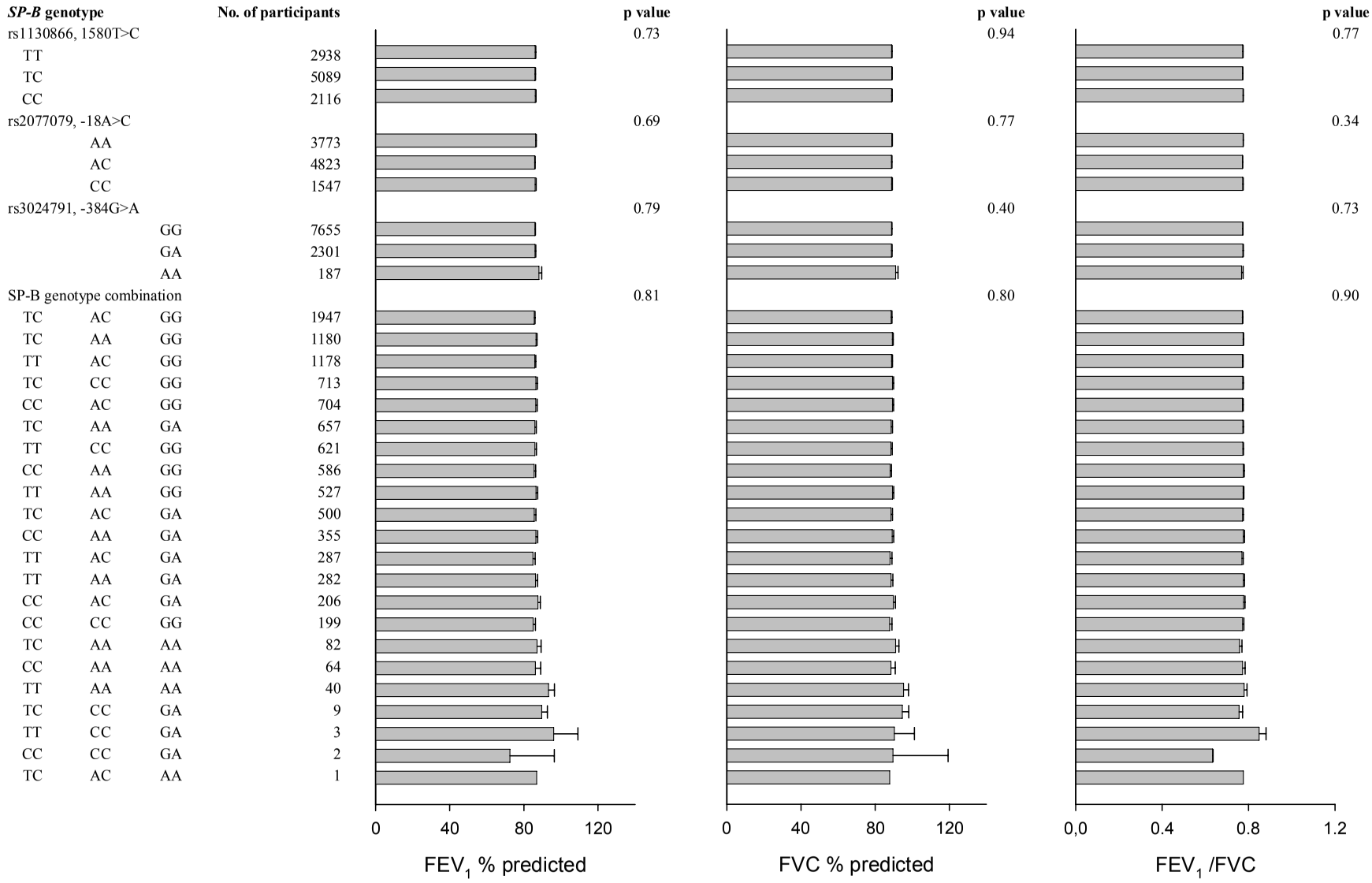

SP-B genotype
No. of participants
p value
p value
p value
rs1130866, 1580T>C
0.73
0.94
0.77
TT 2938
TC 5089
CC 2116
rs2077079, -18A>C
0.69
0.77
0.34
AA 3773
AC 4823
CC 1547
rs3024791, -384G>A
0.79
0.40
0.73
GG 7655
GA 2301
AA 187
SP-B genotype combination
0.81
0.80
0.90
TC AC GG 1947
TC AA GG 1180
TT AC GG 1178
TC CC GG 713
CC AC GG 704
TC AA GA 657
TT CC GG 621
CC AA GG 586
TT AA GG 527
TC AC GA 500
CC AA GA 355
TT AC GA 287
TT AA GA 282
CC AC GA 206
CC CC GG 199
TC AA AA 82
CC AA AA 64
TT AA AA 40
TC CC GA 9
TT CC GA 3
CC CC GA 2
TC AC AA 1
0
40
80
120
FEV1 % predicted
0
40
80
120
FVC % predicted
0,0
0.4
0.8
1.2
FEV1 /FVC

Figure 2

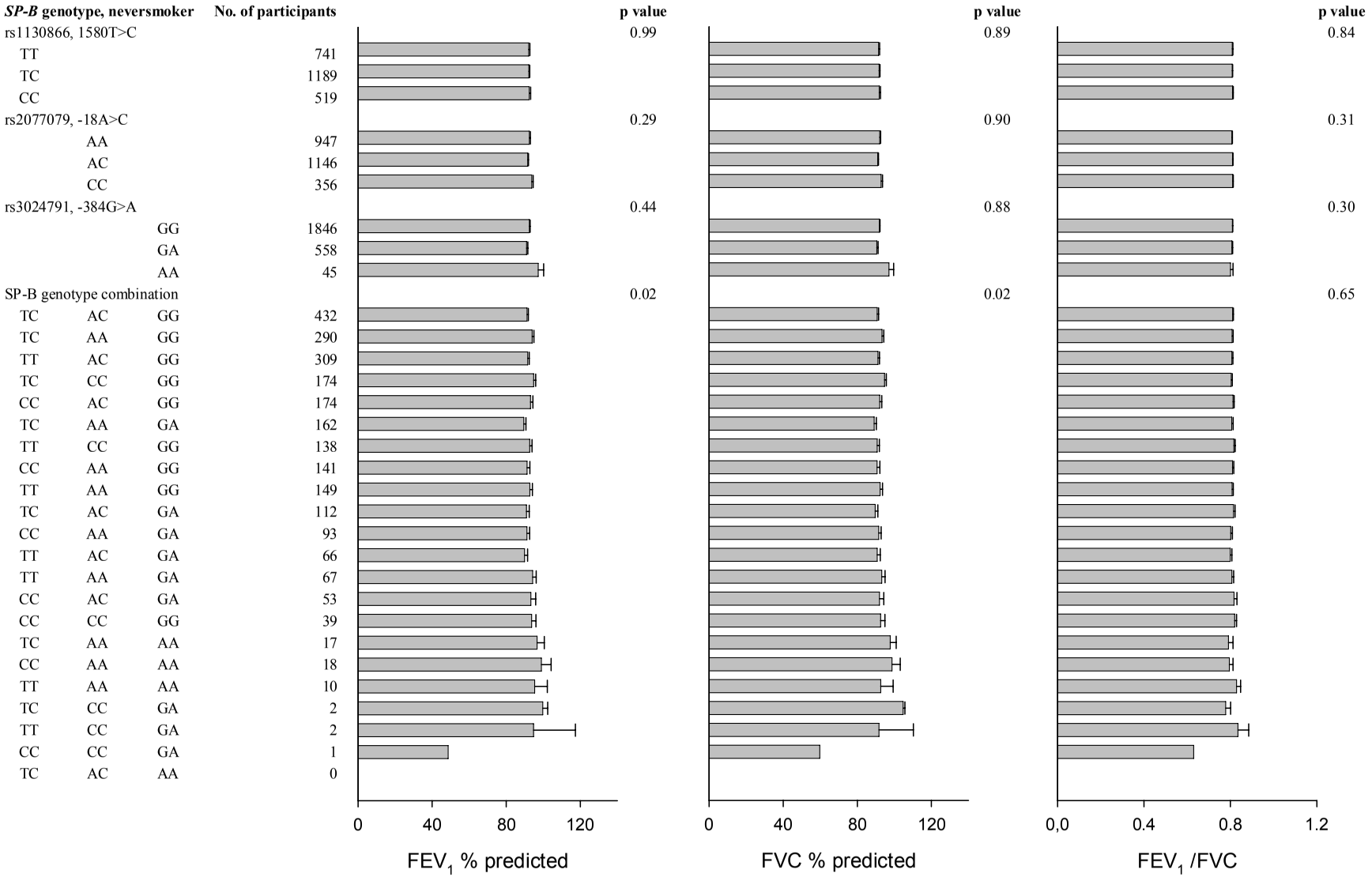

SP-B genotype, neversmoker
No. of participants
p value
p value
p value
rs1130866, 1580T>C
0.99
0.89
0.84
TT 741
TC 1189
CC 519
rs2077079, -18A>C
0.29
0.90
0.31
AA 947
AC 1146
CC 356
rs3024791, -384G>A
0.44
0.88
0.30
GG 1846
GA 558
AA 45
SP-B genotype combination
0.02
0.02
0.65
TC AC GG 432
TC AA GG 290
TT AC GG 309
TC CC GG 174
CC AC GG 174
TC AA GA 162
TT CC GG 138
CC AA GG 141
TT AA GG 149
TC AC GA 112
CC AA GA 93
TT AC GA 66
TT AA GA 67
CC AC GA 53
CC CC GG 39
TC AA AA 17
CC AA AA 18
TT AA AA 10
TC CC GA 2
TT CC GA 2
CC CC GA 1
TC AC AA 0
0 40 80 120
FEV1 % predicted
0 40 80 120
FVC % predicted
0,0 0.4 0.8 1.2
FEV1 /FVC

Figure 3

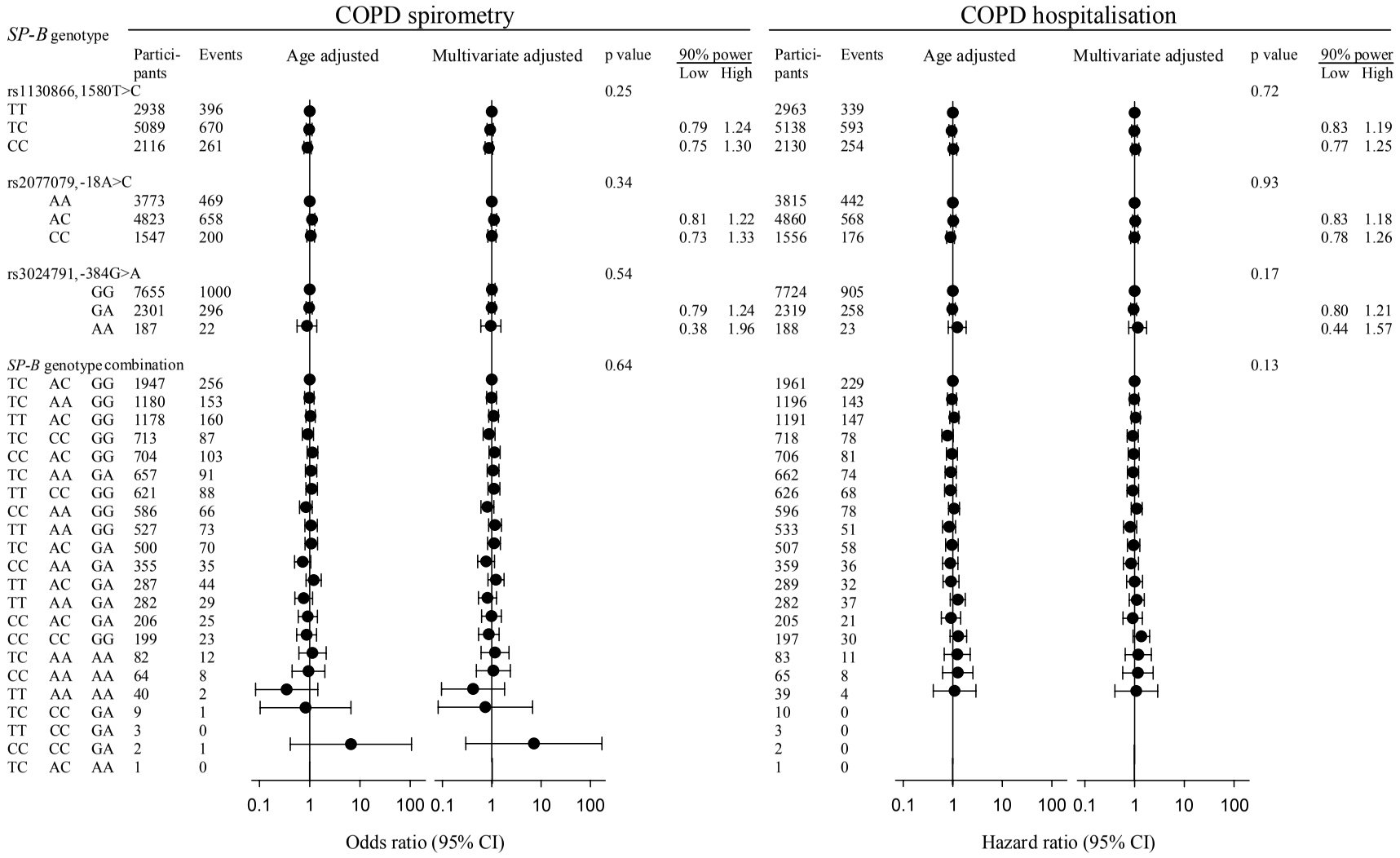
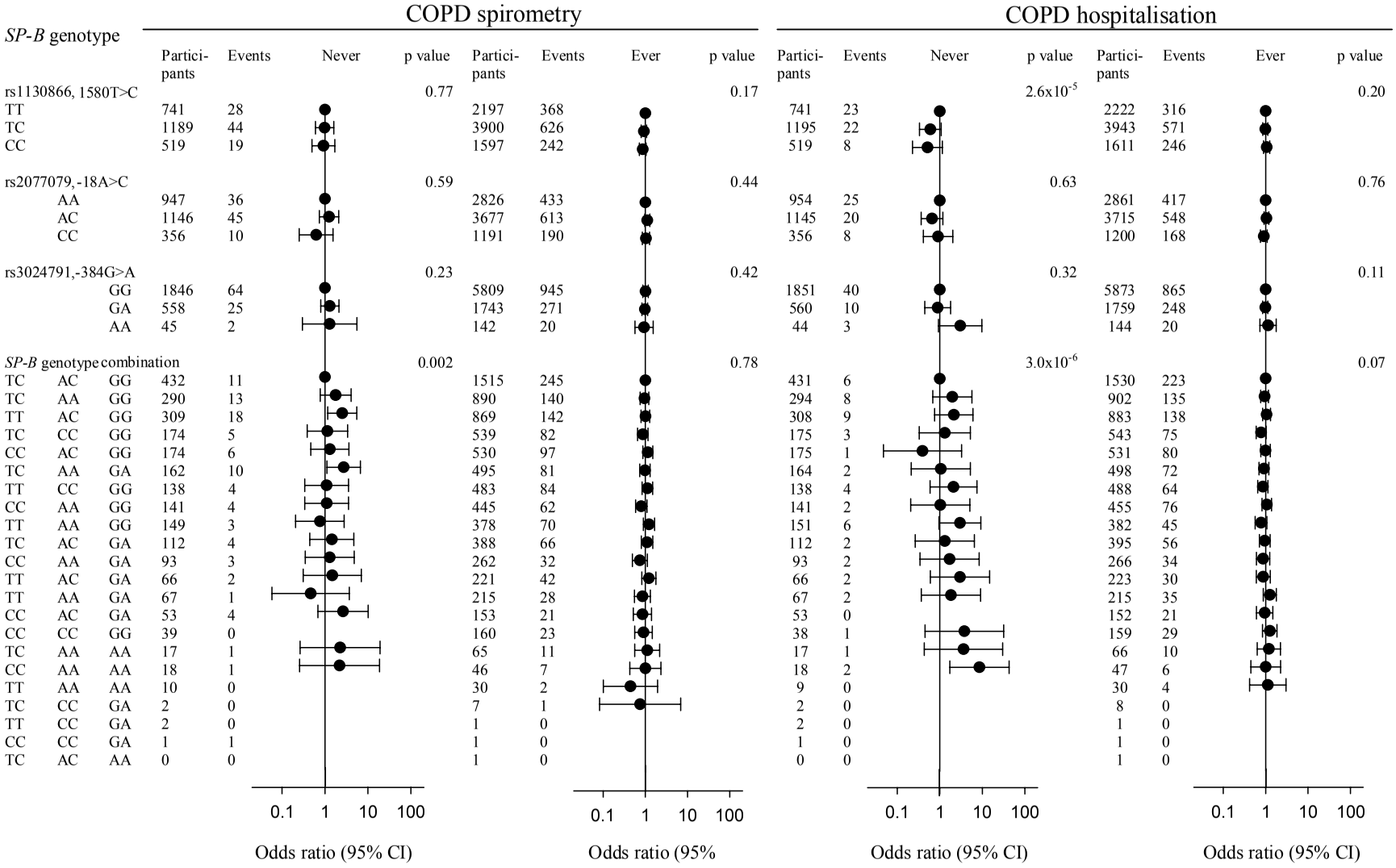

COPD spirometry
COPD hospitalisation
SP-B genotype
Partici-pants
Events
Age adjusted
Multivariate adjusted
p value
90% power
Low
High
rs1130866,1580T>C
0.25
0.72
TT 2938 396 2963 339
TC 5089 670 0.79 1.24 5138 593 0.83 1.19
CC 2116 261 0.75 1.30 2130 254 0.77 1.25
rs2077079,-18A>C
0.34
0.93
AA 3773 469 3815 442
AC 4823 658 0.81 1.22 4860 568 0.83 1.18
CC 1547 200 0.73 1.33 1556 176 0.78 1.26
rs3024791,-384G>A
0.54
0.17
GG 7655 1000 7724 905
GA 2301 296 0.79 1.24 2319 258 0.80 1.21
AA 187 22 0.38 1.96 188 23 0.44 1.57
SP-B genotype combination
0.64
0.13
TC AC GG 1947 256 1961 229
TC AA GG 1180 153 1196 143
TT AC GG 1178 160 1191 147
TC CC GG 713 87 718 78
CC AC GG 704 103 706 81
TC AA GA 657 91 662 74
TT CC GG 621 88 626 68
CC AA GG 586 66 596 78
TT AA GG 527 73 533 51
TC AC GA 500 70 507 58
CC AA GA 355 35 359 36
TT AC GA 287 44 289 32
TT AA GA 282 29 282 37
CC AC GA 206 25 205 21
CC CC GG 199 23 197 30
TC AA AA 82 12 83 11
CC AA AA 64 8 65 8
TT AA AA 40 2 39 4
TC CC GA 9 1 10 0
TT CC GA 3 0 3 0
CC CC GA 2 1 2 0
TC AC AA 1 0 1 0
0.1
1
10
100
Odds ratio (95% CI)
Hazard ratio (95% CI)

Figure 4


COPD spirometry
COPD hospitalisation
SP-B genotype
Never
Ever
p value
rs1130866, 1580T>C
TT 741 28 2197 368
TC 1189 44 3900 626
CC 519 19 1597 242
rs2077079, -18A>C
AA 947 36 2826 433
AC 1146 45 3677 613
CC 356 10 1191 190
rs3024791, -384G>A
GG 1846 64 5809 945
GA 558 25 1743 271
AA 45 2 142 20
SP-B genotype combination
TC AC GG 432 11 1515 245
TC AA GG 290 13 890 140
TT AC GG 309 18 869 142
TC CC GG 174 5 539 82
CC AC GG 174 6 530 97
TC AA GA 162 10 495 81
TT CC GG 138 4 483 84
CC AA GG 141 4 445 62
TT AA GG 149 3 378 70
TC AC GA 112 4 388 66
CC AA GA 93 3 262 32
TT AC GA 66 2 221 42
TT AA GA 67 1 215 28
CC AC GA 53 4 153 21
CC CC GG 39 0 160 23
TC AA AA 17 1 65 11
CC AA AA 18 1 46 7
TT AA AA 10 0 30 2
TC CC GA 2 0 7 1
TT CC GA 2 0 1 0
CC CC GA 1 1 1 0
TC AC AA 0 0 1 0
0.77
0.17
0.59
0.44
0.23
0.42
0.002
0.78
2.6x10-5
0.20
0.63
0.76
0.32
0.11
3.0x10-6
0.07
741 23 2222 316
1195 22 3943 571
519 8 1611 246
954 25 2861 417
1145 20 3715 548
356 8 1200 168
1851 40 5873 865
560 10 1759 248
44 3 144 20
431 6 1530 223
294 8 902 135
308 9 883 138
175 3 543 75
175 1 531 80
164 2 498 72
138 4 488 64
141 2 455 76
151 6 382 45
112 2 395 56
93 2 266 34
66 2 223 30
67 2 215 35
53 0 152 21
38 1 159 29
17 1 66 10
18 2 47 6
9 0 30 4
2 0 8 0
2 0 1 0
1 0 1 0
0 0 1 0
0.1 1 10 100
Odds ratio (95% CI)
Odds ratio (95%
Odds ratio (95% CI)
Odds ratio (95% CI)