A predictive model for the development of chronic obstructive pulmonary disease

YI GUO[1*], YANRONG QIAN[1*], YI GONG[2], CHUNMING PAN[3], GUOCHAO SHI[1] and HUANYING WAN[1]

[1]Department of Pulmonary Medicine, Ruijin Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai 200025;
[2]Department of Pulmonary Medicine, Huashan Hospital, Fudan University, Shanghai 200040;
[3]State Key Laboratory of Medical Genomics, Molecular Medicine Center,
Shanghai Ruijin Hospital, Shanghai 200025, P.R. China

Received December 23, 2014; Accepted February 12, 2015

DOI: 10.3892/br.2015.503

Abstract. The screening of a person at risk for chronic obstructive pulmonary disease (COPD) and timely treatment may provide opportunities to delay the progressive destruction of lung function. Therefore, a model to predict the disease is required. We hypothesized that demographic and clinical information in combination with genetic markers would aid in the prediction of COPD development, prior to its onset. The aim of the present study was to create a predictive model for COPD development. Demographic, clinical presentation and genetic polymorphisms were recorded in COPD patients and control subjects. Nighty-six single-nucleotide polymorphisms of 46 genes were selected for genotyping in the case-control study. A predictive model was produced using logistic regression with a stepwise model-building approach and was validated. A total of 331 patients and 351 control subjects were included. The logistic regression identified the following predictors: Gender, respiratory infection in early life, low birth weight, smoking history and genotype polymorphisms (rs2070600, rs10947233, rs1800629, rs2241712 and rs1205). The model was established using the following formula: COPD = 1/[1 + exp (-2.4933-1.2197 gender + 1.1842 respiratory infection in early life + 2.4350 low birth weight + 1.8524 smoking - 1.1978 rs2070600 + 2.0270 rs10947233 + 1.1913 rs10947233 + 0.6468 rs1800629 + 0.5272 rs2241712 + 0.4024 rs1205)] (when the value is >0.5). The Hosmer-Lemeshow test showed no significant deviations between the observed and predicted events. Validation of the model in 50 patients showed a modest sensitivity and specificity. Therefore, a predictive model based on demographic, clinical and genetic information may identify COPD prior to its onset.

Introduction

Chronic obstructive pulmonary disease (COPD) is characterized by progressive airflow limitation, driven by an abnormal inflammatory response of the airways to inhaled particles and fumes (1). The disease is predicted to become the third most common cause of mortality and the fifth cause of disability in the world by 2020 (2). COPD represents a significant burden for the health care systems worldwide (3).

COPD is also causing an increasing problem in China. A survey conducted in 2007 of 20,245 participants in seven regions of China indicated that the prevalence of COPD in adults aged ≥40 years was 8.2% (4). However, numerous patients with COPD remain undiagnosed until the more advanced stages of the disease. A study by Professor Nanshan Zhong (5), the Chief of the Chinese Medicine Association, showed that the diagnosis was established only in 31% of the COPD patients. A number of population-based studies revealed that the disease was also under-diagnosed in other countries (6-8). In a study of Spanish patients (9), only 25% of smokers with COPD were previously known regarding the diagnosis. Additionally, <50% of patients with severe or extremely severe airflow obstruction were diagnosed (10). COPD is usually diagnosed in the later stage when significant lung function has already been lost, being asymptomatic in the early phase, and sometimes patients are not diagnosed until they are hospitalized for an acute exacerbation (11). However, the airway limitation is much more reversible in early COPD, as early detection and timely treatment can slow the destruction of lung function. Therefore, a predictive model for COPD development that could have a clinical utility is required. Previous studies (12,13) of COPD predictors identified certain risk factors, including age, smoking, forced expiratory volume in 1 sec (FEV1), low body weight and poor performance status, but a single determinant was not reliable to estimate the probability of COPD development, therefore, a full predictive model must be developed using comprehensive indicators.

In addition, the natural history of the development of the disease in smokers is highly variable, as only a minority of smokers (20%) appear to present airflow limitation, suggesting

Correspondence to: Professor Huanying Wan, Department of Pulmonary Medicine, Ruijin Hospital, School of Medicine, Shanghai Jiao Tong University, 197 Ruijin Er Road, Shanghai 200025, P.R. China
E-mail: hy_wan2013@163.com

*Contributed equally

Key words: chronic obstructive pulmonary disease, predictive model, single-nucleotide polymorphism, genotype, risk factors

that besides smoking, COPD is partially genetically determined (14,15). Genes were evidenced to be associated with familial aggregation of COPD (16), and certain other twin studies have also indicated a genetic contribution to clinically relevant parameters on pulmonary function (17). Genome-wide association studies (GWAS) have identified certain susceptibility loci, but these are few in the Asian population (18,19). Consequently, we hypothesized that the abovementioned risk factors in combination with genetic markers would aid the prediction of COPD development prior to its onset.

The aim of the present study was to set up a predictive model for COPD development in a Chinese population. First, the candidate genes for the susceptibility to COPD were identified among 97 single-nucleotide polymorphisms (SNPs) of 46 genes. Second, a mathematical formula based on the clinical and demographic data recorded combined with SNP markers was produced.

Materials and methods

Part I

Study population of SNP identification. A total of 331 unrelated adult patients with COPD were recruited from the Department of Pulmonary Medicine of Shanghai Ruijin Hospital (Shanghai, China) between January 2012 and November 2013. COPD was diagnosed according to the criteria established by the National Heart, Lung and Blood Institute/World Health Organization Global Initiative for COPD (GOLD) (20). The entry criteria were as follows: Presence of relentlessly progressive symptoms, such as cough, productive sputum or breathlessness; age, ≥40 years; airflow limitation as indicated by FEV1/forced vital capacity (FVC) ≤70%; FEV1 reversibility following the inhalation of salbutamol <12% of the pre-bronchodilator FEV1 (MS-Body Diffusion; Jaeger GmbH, Würzburg, Germany); and no evidence of hereditary diseases or other respiratory diseases.

A total of 213 control healthy smokers were selected from a pool of healthy subjects who visited the General Health Checkup Center of Shanghai Ruijin Hospital in the same period. The enrollment criteria for the controls were as follows: Age ≥40 years, smoker, no known disease, no history of any disease and lung function was measured at baseline following the American Thoracic Society/European Respiratory Society standard procedure to confirm no evidence of airflow obstruction. All the cases and control subjects were Chinese. The study protocol was approved by the Medical Ethics Committee of Shanghai Ruijin Hospital and all the participants provided written informed consent.

DNA extraction and genotyping. According to the results of previous GWAS, 97 candidate SNPs were chosen for genotyping (Table I). Their minor allele frequencies were >0.05 in the Chinese patients. A peripheral blood sample was obtained from each participant and DNA was isolated using QuickGene DNA Whole Blood kit (Fujifilm Life Science, Tokyo, Japan). Any sample with a DNA concentration <10 ng/μl was excluded and required another sample. The Mass-Array™ Technology platform of Sequenom, Inc., (San Diego, CA, USA) was used to perform genotyping. For quality control, two independent investigators interpreted the results and a random selection of 10% of all the samples was re-tested. Each of the SNPs in the control group was analyzed for the Hardy-Weinberg equilibrium (HWE), and SNPs were excluded from the analysis if they were out of HWE (P≤0.05). The χ^2 test and unconditional logistic method were applied to compare the allele frequencies between the two groups, and logistic analysis was adjusted for age, gender and smoking. Frequencies were compared, respectively, using a P cut-off of 0.05 and the Bonferroni correction method for multiple testing in order to identify several SNPs in susceptibility to COPD. P<0.05 was considered to indicate a statistically significant difference.

Part II

Study population of predictive model-building. In total, 331 COPD patients and 351 control subjects were recruited from the Department of Pulmonary Medicine between January 2012 and December 2013. All the patients met the diagnostic criteria of GOLD and were ≥40 years. The control subjects were present with no evidence of airflow obstruction, aged ≥40 years, and were smokers or non-smokers. They had no hereditary diseases or other respiratory diseases.

SNP genotyping. A peripheral blood sample was obtained from each participant and DNA was isolated using the same methods, as previously described. The SNPs identified in the susceptibility to COPD in part I were genotyped.

Documentation of data. In addition to the SNP genotyping, demographic data, body mass index, history of respiratory infection in childhood, low birth weight (<2,500 g), environmental pollution (their place of residence and work environment), smoking history, family history of lung disease, and spirometry of these 682 subjects were recorded. The case group was defined as 1, the control group as 0; similarly, 1=male, 0=female; 1=respiratory infection in childhood, 0=no infection; 1=history of low birth weight, 0=non low birth weight; 1=environmental pollution, 0=no exposure; 1=smoking history, 0=non smoking; and 1=family history of lung disease, 0=no known family history. These risk factors were identified in association to COPD based on our previous epidemiology study (21). Genotyping results were also recorded using 0 or 1.

Predictive model-building methods. The predictive model was constructed by means of logistic regression with a stepwise model-building approach, using an entry and exit criterion of P≤0.05. The variables included genetic polymorphisms verified according to the results of genotyping and clinical data of each participant recorded above. The goodness of fit, namely how closely the prediction reflected observed events, was determined by the Hosmer-Lemeshow test.

Statistical analysis. Data analyses were performed with the Statistical Package for the Social Science version 20.0 (SPSS, Inc., Chicago, IL, USA) and P<0.05 was considered to indicate a statistically significant difference. The two-sided Student's t-test was used for checking the significant differences in the clinical data between the cases and control subjects. The relative risk of the allelic gene was estimated as an odds ratio with a 95% confidence interval.

Results

Part I

Study population characteristics. The study population characteristics are described in Table II. They were matched for gender and age. FEV1 predictive and FEV1/FVC of the case

Table I. Gene location and alleles of 97 single-nucleotide polymorphisms (SNPs).

SNP_ID (Refs.)	Gene	Chromosome	Alleles	SNP_ID (Refs.)	Gene	Chromosome	Alleles
rs1800610 (1)	*TNF-α*	6	C/T	rs673400 (14)	*SERPINA2*	2	C/G
rs1799964 (1)	*TNF-α*	6	C/T	rs7583463 (15)	*SERPINA2*	2	A/C
rs361525 (2)	*TNF-α*	6	A/G	rs2736100 (8)	*TERT*	5	G/T
rs1800629 (3)	*TNF-α*	6	A/G	rs10069690 (8)	*TERT*	5	C/T
rs2808630 (4)	*CRP*	1	C/T	rs34829399 (8)	*TERT*	5	C/T
rs1205 (5)	*CRP*	1	C/T	rs4246742 (8)	*TERT*	5	A/T
rs1130864 (4)	*CRP*	1	C/T	rs2736118 (8)	*TERT*	5	A/G
rs1059823 (6)	*SLC11A1*	2	A/G	rs2736122 (8)	*TERT*	5	C/T
rs1130866 (7)	*SFTPB*	2	C/T	rs2853677 (8)	*TERT*	5	C/T
rs2353397 (8)	*HHIP*	4	C/T	rs2853676 (8)	*TERT*	5	A/G
rs13147758 (8)	*HHIP*	4	A/G	rs1881457 (16)	*IL-13*	5	A/C
rs2035901 (8)	*HHIP*	4	A/G	rs1295685 (16)	*IL-13*	5	C/T
rs6537302 (8)	*HHIP*	4	A/T	rs1800925 (16)	*IL-13*	5	C/T
rs1032295 (8)	*HHIP*	4	T/G	rs2066960 (16)	*IL-13*	5	A/C
rs12504628 (8)	*HHIP*	4	C/T	rs20541 (16)	*IL-13*	5	C/T
rs17019336 (8)	*HHIP*	4	A/T	rs16909898 (8)	*PTCH1*	9	A/G
rs3749893 (8)	*TSPYL-4*	6	A/G	rs10512249 (8)	*PTCH1*	9	C/T
rs4987835 (9)	*Bcl-2*	18	A/G	rs35621 (17)	*ABCC1*	16	C/T
rs2292566 (10)	*EPHX1*	1	A/G	rs2241718 (18)	*TGF-β1*	19	C/T
rs1051740 (11)	*EPHX1*	1	C/T	rs56155294 (18)	*TGF-β1*	19	C/T
rs868966 (11)	*EPHX1*	1	A/G	rs1800469 (18)	*TGF-β1*	19	C/T
rs25882 (12)	*CSF2*	5	C/T	rs2241712 (18)	*TGF-β1*	19	A/G
rs829259 (13)	*PDE4D*	5	A/T	rs2277027 (8)	*ADAM19*	5	A/C
rs6712954 (14)	*SERPINA2*	2	A/G	rs2280090 (19)	*ADAM33*	20	A/G
rs2280091(19)	*ADAM33*	20	A/G	rs4073 (12)	*IL-8*	4	A/T
rs1435867 (8)	*PID1*	2	C/T	rs8192288 (30)	*SOD3*	4	G/T
rs10498230 (8)	*PID1*	2	C/T	rs2571445 (20)	*TNS1*	2	C/T
rs3995090 (20)	*HTR4*	5	A/C	rs1003349 (31)	*MMP14*	14	G/T
rs6889822 (8)	*HTR4*	5	A/G	rs737693 (32)	*MMP12*	11	A/T
rs1531697 (9)	*Bcl-2*	18	A/T	rs2276109 (32)	*MMP12*	11	A/G
rs1042713 (21)	*ARDB2*	5	A/G	rs1052443 (8)	*NT5DC1*	6	A/C
rs3024791 (22)	*SFTPB*	2	A/G	rs10947233 (8)	*PPT2*	6	G/T
rs511898 (23)	*ADAM33*	20	C/T	rs1051730 (33)	*CHRNA3*	15	C/T
rs2853209 (23)	*ADAM33*	20	A/T	rs11106030 (20)	*DCN*	12	A/C
rs6555465 (8)	*ADCY2*	5	A/G	rs584367 (34)	*sPLA2s*	1	C/T
rs10075508 (13)	*PDE4D*	5	C/T	rs9904270 (26)	*CDC6*	17	C/T
rs12899618 (20)	*THSD4*	15	A/G	rs2395730 (8)	*DAAM2*	6	A/C
rs3091244 (8)	*SFXN1*	5	A/C/T	rs3817928 (8)	*GPR126*	6	A/G
rs8004738 (24)	*SERPINA1*	14	A/G	rs11155242 (8)	*GRP126*	6	A/C
rs709932 (24)	*SERPINA1*	14	A/G	rs7776375 (8)	*GPR126*	6	A/G
rs4934 (25)	*SERPINA3*	14	A/G	rs6937121 (8)	*GPR126*	6	G/T
rs13706 (26)	*CDC6*	17	A/G	rs1042714 (35)	*ARDB2*	5	C/G
rs7217852 (26)	*CDC6*	17	A/G	rs1800796 (36)	*IL-6*	7	C/G
rs2077464 (26)	*CDC6*	17	A/G	rs2236307 (31)	*MMP14*	14	C/T
rs2070600 (20)	*AGER*	6	A/G	rs2236302 (31)	*MMP14*	14	C/G
rs6957 (27)	*CDC97*	19	A/G	rs2230054 (37)	*IL-8RB*	2	C/T
rs1042522 (28)	*P53*	17	C/G	rs1422795 (8)	*ADAM19*	5	A/G
rs1695 (29)	*GSTP1*	11	A/G	rs6830970 (8)	*FAM13A*	4	A/G
rs2869967 (8)	*FAM13A*	4	C/T				

group decreased significantly compared to the control group (P<0.05).

Univariate analysis of each genotype. Eight SNPs with a deviation from HWE in the controls were removed from the association analysis; rs361525, rs1042713, rs34829399, rs2853677, rs2571445, rs8192288, rs2066960 and rs2230054. Thirteen SNPs (rs1130866, rs56155294, rs10498230, rs2035901, rs3091244, rs511898, rs2869967, rs7583463, rs2276109, rs737693, rs9904270, rs4934 and rs6830970) were also eliminated for missing data of genotyping in ≥10% of samples. Finally,

Table II. Demographics of COPD patients and control subjects.

Variables	COPD	Controls	P-value
Subjects, n	331	213	
Age, years	61±10	58±12	
Male, n (%)	298 (90)	209 (98)	
Female, n (%)	33 (10)	4 (2)	
Pack-years	41±34	38±17	
FEV1/FVC	54±13.8[a]	85±7.6	<0.05
FEV1 predicted, %	49±18.1[a]	88±17.0	<0.05

[a]P<0.05, verses control. Data are presented as the means ± standard deviation. COPD, chronic obstructive pulmonary disease; FEV1, forced expiratory volume in 1 sec; FVC, forced vital capacity.

76 of the 97 SNPs were included in the association analysis. The allele frequencies and the genotype distributions for these SNPs were compared between the patients and control healthy smokers. Several allelic genes of seven SNPs were found to be more frequent in the COPD patients compared to the control subjects. These were human hedgehog interacting protein (*HHIP*) (rs2353397 C allele) (P<0.0001), *TNF-α* (rs1800629 G allele) (P=0.0060), *TGF-β1* (rs2241712 A allele) (P=0.0498), *CRP* (rs1205 C allele) (P=0.0030), *IL-13* (rs20541 T allele) (P=0.0280), *AGER* (rs2070600 G allele) (P=0.0130) and *PPT2* (rs10947233 G allele) (P=0.0060). These seven SNPs tended to be associated with COPD. Among these seven SNPs, following Bonferroni correction, rs2353397 (P<0.0001) was most strongly associated with the susceptibility to COPD (Table III).

Part II

Predictive model for COPD. The clinical data of the 331 COPD patients and 351 control subjects recruited for the second part of the study were recorded. Clinical variables recorded for the logistic regression model are presented in Table IV. The genotype of the seven SNPs was also recorded. Genetic variables that achieved significance in univariate analysis were defined as follows: CT=1 0, TT=0 0, CC=0 1 (rs2353397); GA=1 0, AA=0 0, GG=0 1 (rs2070600); GT=1 0, TT=0 0, GG=0 1 (rs10947233); GA=1 0, AA=0 0, GG=0 1 (rs1800629); AG=1 0, GG=0 0, AA=0 1 (rs2241712); CT=1 0, TT=0 0, CC=0 1 (rs1205); and TC=1 0, CC=0 0, TT=0 1 (rs20541). The different genotypes combined with the clinical data of the two groups were entered in the multivariate analysis, which was performed using the logistic regression model. Finally, the model was established using the following formula: (P-value for

Table III. Allele frequencies in COPD and control subjects for SNPs.

SNP	Allele	Control, n (%)	Case, n (%)	χ^2	P-value	OR (95% CI)	$P_{(Bonferroni)}$	Adjusted P-value	Adjusted OR (95% CI)	Adjusted $P_{(Bonferroni)}$
rs1059823	G	139 (33)	222 (34)	0.0181	0.8929	1.01 (0.79-1.32)	67.8604	0.8290	0.97 (0.74-1.27)	63.0040
	A	283 (67)	440 (66)							
rs1205	C	168 (40)	308 (47)	5.2168	0.0223[a]	1.34 (1.04-1.71)	1.6948	0.0030[a]	1.48 (1.14-1.91)	0.2280
	T	252 (60)	346 (53)							
rs17019336	A	136 (32)	242(37)	2.1770	0.1401	1.21 (0.94-1.57)	10.6476	0.0670	1.28 (0.98-1.68)	5.0920
	T	284 (68)	416 (63)							
rs1799964	T	333 (79)	519 (79)	0.0008	0.9772	1.00 (0.74-1.36)	74.2672	0.8140	0.96 (0.71-1.31)	61.8640
	C	87 (21)	135 (21)							
rs1800610	T	71 (17)	112 (17)	0.0117	0.9137	1.02 (0.74-1.41)	69.4412	0.9007	0.98 (0.70-1.38)	68.4532
	C	355 (83)	550 (83)							
rs2077464	T	271 (65)	420 (66)	0.1909	0.6621	1.06 (0.82-1.37)	50.3196	0.8230	1.03 (0.79-1.35)	62.5480
	C	149 (35)	218 (34)							
rs2236302	C	369 (88)	584 (89)	0.2011	0.6539	1.09 (0.75-1.59)	49.6964	0.4140	1.18 (0.80-1.75)	31.4640
	G	51 (12)	74 (11)							
rs2292566	A	125 (30)	209 (32)	0.4800	0.4884	1.10 (0.84-1.43)	37.1184	0.7630	1.04 (0.79-1.38)	57.9880
	G	295 (70)	449 (68)							
rs2353397	C	123 (29)	382 (58)	83.3798	$6.8x10^{-20a}$	3.29 (2.54-4.28)	$5.2x10^{-18a}$	<0.0001[a]	2.16 (1.66-2.81)	<0.0001[a]
	T	297 (71)	280 (42)							
rs25882	T	147 (35)	240 (36)	0.2421	0.6227	1.07 (0.83-1.38)	47.3252	0.4650	1.10 (0.85-1.44)	35.3400
	C	273 (65)	418 (64)							
rs2808630	C	66 (16)	119 (18)	1.0136	0.3140	1.18 (0.85-1.65)	23.8640	0.2120	0.86 (0.69-1.09)	16.1120
	T	354 (84)	539 (82)							
rs3749893	A	286 (67)	454 (69)	0.6232	0.4299	1.11 (0.86-1.44)	32.6724	0.4510	1.11 (0.84-1.46)	34.2760
	G	140 (33)	200 (31)							
s4987835	A	236 (56)	382 (60)	1.7185	0.1899	1.19 (0.92-1.51)	14.4324	0.2950	1.15 (0.88-1.49)	22.4200
	G	184 (44)	252 (40)							
rs709932	A	73 (17)	131 (20)	1.2714	0.2595	1.20 (0.87-1.65)	19.7220	0.2860	1.19 (0.86-1.65)	21.7360
	G	347 (83)	519 (80)							

Table III. Continued.

SNP	Allele	Control, n (%)	Case, n (%)	χ^2	P-value	OR (95% CI)	$P_{(Bonferroni)}$	Adjusted P-value	Adjusted OR (95% CI)	Adjusted $P_{(Bonferroni)}$
rrs7217852	A	273 (65)	434 (66)	0.0652	0.7985	1.03 (0.80-1.34)	60.6860	0.8460	1.03 (0.79-1.34)	64.2960
	G	147 (35)	226 (34)							
rs7776375	A	270 (63)	438 (66)	0.8832	0.3473	1.13 (0.88-1.46)	26.3948	0.2570	1.17 (0.89-1.52)	19.5320
	G	156 (37)	224 (34)							
rs10069690	C	331 (80)	520 (81)	0.0264	0.8709	1.03 (0.75-1.40)	66.1884	0.6480	1.08 (0.78-1.48)	49.2480
	T	81 (20)	124 (19)							
rs1051740	T	247 (60)	403 (61)	0.0424	0.8369	1.03 (0.79-1.32)	63.6044	0.8910	1.02 (0.79-1.32)	67.7160
	C	163 (40)	259 (39)							
rs11155242	A	372 (90)	604 (91)	0.5784	0.4469	1.18 (0.77-1.79)	33.9644	0.2560	1.28 (0.83-1.94)	19.4560
	C	42 (10)	58 (9)							
rs1295685	T	118 (29)	221 (33)	2.8124	0.0935	1.26 (0.96-1.64)	7.1060	0.1730	1.21 (0.92-1.60)	13.1480
	C	296 (71)	441 (67)							
rs1435867	C	55 (13)	90 (14)	0.0200	0.8877	1.03 (0.72-1.47)	67.4652	0.5300	0.89 (0.62-1.29)	40.5080
	T	355 (87)	566 (86)							
rs16909898	G	33 (8)	54 (8)	0.0101	0.9200	1.02 (0.65-1.61)	69.9200	0.3140	0.79 (0.50-1.25)	23.8640
	A	379 (92)	606 (92)							
rs1881457	A	308 (74)	495 (75)	0.0726	0.7876	1.04 (0.78-1.38)	59.8576	0.9120	1.02 (0.76-1.36)	69.3120
	C	108 (26)	167 (25)							
rs2241718	T	114 (28)	206 (31)	1.4433	0.2296	1.18 (0.90-1.55)	17.4496	0.2930	1.16 (0.88-1.53)	22.2680
	C	298 (72)	456 (69)							
rs2277027	C	64 (15)	106 (16)	0.0586	0.8088	1.04 (0.74-1.46)	61.4688	0.8350	0.96 (0.68-1.36)	63.4600
	A	350 (85)	556 (84)							
rs2736100	T	231 (57)	368 (58)	0.1539	0.6948	1.05 (0.82-1.35)	52.8048	0.6340	1.06 (0.82-1.38)	48.1840
	G	173 (43)	262 (42)							
rs35621	C	305 (74)	499 (75)	0.1317	0.7167	1.05 (0.79-1.40)	54.4692	0.3480	1.15 (0.86-1.54)	26.4480
	T	105 (26)	163 (25)							
rs3995090	C	288 (70)	461 (71)	0.1521	0.6965	1.06 (0.80-1.39)	52.9340	0.4200	1.12 (0.84-1.48)	31.9200
	A	122 (30)	185 (29)							
rs4246742	A	244 (60)	429 (65)	3.0339	0.0815	1.25 (0.97-1.61)	6.1940	0.0510	1.32 (1.01-1.71)	3.8760
	T	166 (40)	233 (35)							
rs6712954	G	321 (78)	545 (82)	3.1679	0.0751	1.32 (0.97-1.79)	5.7076	0.0560	1.38 (1.01-1.89)	4.2560
	A	91 (22)	117 (18)							
rs829259	A	137 (33)	233 (35)	0.4250	0.5145	1.09 (0.84-1.41)	39.1020	0.9300	1.01 (0.77-1.32)	70.6800
	T	275 (67)	429 (65)							
rs10075508	T	69 (16)	108 (17)	0.0253	0.8736	1.03 (0.74-1.43)	66.3936	0.9070	1.02 (0.72-1.44)	68.9320
	C	357 (84)	544 (83)							
rs10512249	T	33 (8)	52 (8)	0.0606	0.8056	1.06 (0.67-1.67)	61.2256	0.4950	1.16 (0.75-1.80)	37.6200
	C	383 (92)	570 (92)							
rs12899618	G	370 (89)	579 (89)	0.0806	0.7765	1.06 (0.72-1.56)	59.0140	0.6010	1.11 (0.75-1.65)	45.6760
	A	48 (11)	71 (11)							
rs13706	G	272 (65)	427 (65)	0.0598	0.8068	1.03 (0.80-1.34)	61.3168	0.8300	0.97 (0.75-1.27)	63.0800
	A	148 (35)	225 (35)							
rs1531697	A	255 (61)	411 (63)	0.5370	0.4637	1.10 (0.85-1.41)	35.2412	0.4750	1.10 (0.85-1.43)	36.1000
	T	163 (39)	239 (37)							
rs1800925	T	62 (15)	105 (17)	0.8620	0.3531	1.18 (0.84-1.66)	26.8356	0.1000	1.32 (0.94-1.85)	7.6000
	C	352 (85)	507 (83)							
rs3024791	G	388 (93)	616 (95)	1.5299	0.2161	1.39 (0.82-2.34)	16.4236	0.3820	1.25 (0.76-2.06)	29.0320
	A	28 (7)	32 (5)							
rs6537302	A	310 (75)	480 (77)	0.7206	0.3959	1.13 (0.85-1.51)	30.0884	0.9110	1.10 (0.76-1.36)	69.2360
	T	104 (25)	142 (23)							
rs6555465	G	195 (46)	310 (48)	0.4399	0.5072	1.09 (0.85-1.39)	38.5472	0.5010	1.09 (0.85-1.41)	38.0760
	A	231 (54)	338 (52)							
rs673400	C	178 (43)	278 (43)	0.0062	0.9371	1.01 (0.79-1.30)	71.2196	0.9280	0.99 (0.76-1.28)	70.5280
	G	238 (57)	368 (57)							

Table III. Continued.

SNP	Allele	Control, n (%)	Case, n (%)	χ^2	P-value	OR (95% CI)	$P_{(Bonferroni)}$	Adjusted P-value	Adjusted OR (95% CI)	Adjusted $P_{(Bonferroni)}$
rs6889822	G	268 (64)	417 (65)	0.0594	0.8073	1.03 (0.80-1.34)	61.3548	0.5000	1.10 (0.84-1.43)	38.0000
	A	148 (36)	223 (35)							
rs8004738	G	184 (44)	275 (44)	0.0092	0.9236	1.01 (0.79-1.30)	70.1936	0.6650	1.01 (0.82-1.37)	50.5400
	A	232 (56)	351 (56)							
rs1003349	G	238 (57)	392 (60)	0.8897	0.3456	1.13 (0.88-1.45)	26.2656	0.2340	1.17 (0.90-1.51)	17.7840
	T	178 (43)	260 (40)							
rs1032295	T	320 (75)	523 (80)	3.1870	0.0742	1.30 (0.97-1.74)	5.6392	0.1130	1.28 (0.94-1.73)	8.5880
	G	106 (25)	133 (20)							
rs1042522	C	184 (44)	304 (47)	0.8170	0.3660	1.12 (0.88-1.43)	27.8160	0.4090	1.11 (0.86-1.44)	31.0840
	G	236 (56)	348 (53)							
rs1052443	C	281 (67)	457 (71)	1.7602	0.1846	1.20 (0.92-1.56)	14.0296	0.1610	1.22 (0.93-1.60)	12.2360
	A	139 (33)	189 (29)							
rs12504628	T	305 (72)	475 (72)	0.0847	0.7710	1.04 (0.79-1.37)	58.5960	0.9810	1.04 (0.76-1.33)	74.5560
	C	121 (28)	181 (28)							
rs1695	G	72 (17)	126 (19)	0.6673	0.4140	1.14 (0.83-1.57)	31.4640	0.4650	1.13 (0.82-1.57)	35.3400
	A	346 (83)	530 (81)							
rs1800469	C	182 (44)	315 (48)	2.1252	0.1449	1.20 (0.94-1.54)	11.0124	0.2010	1.74 (1.35-2.27)	15.2760
	T	234 (56)	337 (52)							
rs20541	T	118 (28)	228 (35)	5.3633	0.0206[a]	1.37 (1.05-1.79)	1.5656	0.0280[a]	1.36 (1.04-1.80)	2.1280
	C	302 (72)	426 (65)							
rs2070600	G	312 (73)	529 (81)	8.1712	0.0043[a]	1.52 (1.14-2.03)	0.3268	0.0130[a]	1.47 (1.08-1.98)	0.9880
	A	114 (27)	127 (19)							
rs2853209	A	191 (45)	305 (47)	0.1953	0.6586	1.06 (0.83-1.35)	50.0536	0.9890	0.10 (0.77-1.29)	75.1640
	T	231 (55)	349 (53)							
rs4073	A	185 (44)	300 (46)	0.5198	0.4709	1.10 (0.86-1.40)	35.7884	0.2530	1.16 (0.90-1.50)	19.2280
	T	235 (56)	348 (54)							
rs6937121	T	254 (60)	423 (65)	2.1263	0.1448	1.21 (0.94-1.56)	11.0048	0.1720	1.20 (0.92-1.56)	13.0720
	G	166 (40)	229 (35)							
rs6957	G	150 (36)	241 (37)	0.0802	0.7771	1.04 (0.80-1.34)	59.0596	0.6830	1.06 (0.81-1.38)	51.9080
	A	268 (64)	415 (63)							
rs1051730	C	403 (97)	641 (97)	0.2343	0.6284	1.20 (0.57-2.52)	47.3252	0.6480	1.17 (0.60-2.29)	49.2480
	T	11 (3)	21 (3)							
rs10947233	G	299 (72)	526 (79)	7.4524	0.0063[a]	1.49 (1.12-1.98)	0.4788	0.0060[a]	1.51 (1.12-2.03)	0.4560
	T	115 (28)	136 (21)							
rs11106030	C	355 (85)	560 (85)	0.0013	0.9716	1.01 (0.71-1.42)	73.8416	0.7030	1.07 (0.75-1.52)	53.4280
	A	63 (15)	100 (15)							
rs1130864	T	23 (6)	43 (7)	0.6081	0.4355	1.23 (0.73-2.07)	33.0980	0.3890	1.24 (0.77-2.00)	29.5640
	C	389 (94)	591 (93)							
rs1800629	G	379 (90)	627 (95)	7.8793	0.0050[a]	1.94 (1.21-3.10)	0.3800	0.0060[a]	1.97 (1.21-3.21)	0.4560
	A	41 (10)	35 (5)							
rs2241712	A	188 (45)	342 (52)	3.9820	0.0460[a]	1.28 (1.00-1.64)	3.4960	0.0498[a]	1.24 (0.96-1.59)	3.7848
	G	226 (55)	320 (48)							
rs2280090	G	395 (94)	629 (95)	0.9844	0.3211	1.30 (0.77-2.20)	24.4036	0.4640	1.22 (0.72-2.06)	35.2640
	A	27 (6)	33 (5)							
rs2395730	A	119 (28)	209 (32)	1.3886	0.2386	1.17 (0.90-1.54)	18.1336	0.0850	1.28 (0.97-1.69)	6.4600
	C	303 (72)	453 (68)							
rs2736118	A	397 (94)	630 (95)	0.6150	0.4329	1.24 (0.72-2.12)	32.9004	0.2850	1.36 (0.78-2.37)	21.6600
	G	25 (6)	32 (5)							
rs2736122	C	388 (94)	632 (95)	1.5783	0.2090	1.41 (0.82-2.42)	15.8840	0.0510	1.77 (1.02-3.07)	3.8760
	T	26 (6)	30 (5)							
rs3817928	A	370 (89)	596 (90)	0.1202	0.7288	1.07 (0.72-1.61)	55.3888	0.4410	1.17 (0.78-1.76)	33.5160
	G	44 (11)	66 (10)							
rs584367	T	91 (22)	152 (23)	0.1399	0.7083	1.06 (0.79-1.42)	53.8308	0.8590	1.03 (0.76-1.39)	65.2840
	C	323 (78)	510 (77)							

Table III. Continued.

SNP	Allele	Control, n (%)	Case, n (%)	χ^2	P-value	OR (95% CI)	$P_{(Bonferroni)}$	Adjusted P-value	Adjusted OR (95% CI)	Adjusted $P_{(Bonferroni)}$
rs1042714	C	374 (90)	607 (92)	1.1947	0.2744	1.27 (0.83-1.96)	20.8544	0.1440	1.39 (0.90-2.14)	10.9440
	G	40 (10)	374 (90)							
rs13147758	A	283 (69)	464 (71)	0.8780	0.3487	1.14 (0.87-1.49)	26.5012	0.3840	1.13 (0.86-1.49)	29.1840
	G	129 (31)	186 (29)							
rs1422795	G	61 (15)	108 (16)	0.5395	0.4626	1.14 (0.81-1.60)	35.1576	0.8690	1.03 (0.73-1.46)	66.0440
	A	353 (85)	550 (84)							
rs1800796	C	293 (71)	473 (72)	0.1539	0.6948	1.06 (0.80-1.39)	52.8048	0.8250	1.03 (0.78-1.36)	62.7000
	G	121 (29)	185 (28)							
rs2236307	C	169 (41)	286 (43)	0.7270	0.3938	1.11 (0.87-1.43)	29.9288	0.4150	1.11 (0.86-1.44)	31.5400
	T	245 (59)	372 (57)							
rs2280091	A	383 (93)	611 (93)	0.1518	0.6968	1.10 (0.68-1.77)	52.9568	0.5020	1.17 (0.74-1.87)	38.1520
	G	31 (7)	45 (7)							
rs2853676	G	335 (81)	544 (83)	0.4538	0.5005	1.12 (0.81-1.54)	38.0380	0.2770	1.20 (0.86-1.67)	21.0520
	A	77 (19)	112 (17)							
rs868966	A	205 (50)	337 (51)	0.2934	0.5880	1.07 (0.84-1.37)	44.6880	0.7890	1.04 (0.80-1.34)	59.9640
	G	209 (50)	321 (49)							

[a]P<0.05, significant difference is for the alleles between COPD and controls. χ^2 test and logistic analysis were used. Logistic analysis was adjusted by potential confounders, including age, gender and smoking history. COPD, chronic obstructive pulmonary disease; OR, odds ratio; CI, confidence interval.

each variable in Table IV) COPD = 1/[1 + exp (-2.4933-1.2197 gender + 1.1842 respiratory infection in early life + 2.4350 low birth weight + 1.8524 smoking - 1.1978 rs2070600 + 2.0270 rs10947233 + 1.1913 rs10947233 + 0.6468 rs1800629 + 0.5272 rs2241712 + 0.4024 rs1205)] (when the value is >0.5). For example, if the value calculated using the formula above is >0.5 for an individual, it can be speculated that the patient is more likely to develop COPD prior to becoming symptomatic.

Validation of the model. The Hosmer-Lemeshow test showed no significant deviation between the observed and predicted events, suggesting an excellent goodness of fit. Table V shows the results of the test (χ^2=3.948, P=0.862). Data of gender, history of early life respiratory infection, low birth weight, smoking and SNPs identified by logistic regression of 30 COPD patients and 20 healthy controls were entered into the formula, and the values calculated were compared to the observed status. In total, 25 patients obtained values >0.5, and 17 healthy controls had values <0.5 (Table VI). The sensitivity was 83%, specificity was 85%, false negative was 16%, false positive was 15% and Youden index was 0.68.

Discussion

In the present case-control study of 682 participants whose pulmonary function spanned a broad spectrum, a predictive model for development of COPD with a modest sensitivity and specificity was constructed by incorporating demographic, clinical and genetic information, and the statistical model fitted well with the set of observations by the Hosmer-Lemeshow test. The study suggests that the mathematic formula may serve as a helpful tool to identify persons at risk for COPD prior to the onset of symptoms.

Screening for early disease is extremely important, as current medication can only relieve symptoms of COPD, and it has little effect on the delay of its natural progression. Only the person at risk is prospectively identified. Therefore, whether preventive measures can be taken to provide important opportunities for curbing the progressive nature of the disease requires confirmation. Early detection of COPD and intervention for smoking cessation is suggested to delay lung function decline, to reduce the burden of symptoms and to improve the patient quality of life (22,23). However, initially there are no evident symptoms, which becomes a barrier to detection. Therefore, determining how COPD can be detected in the early phase or prior to its onset is required. Given the low diagnostic rate in early phase, the risk assessment for development appears to be valuable. The accurate prediction of the course of airway inflammation in healthy smokers or non-smokers remains a significant challenge.

Thus far, certain studies have focused on identifying tools to diagnose COPD in its earliest stage, but to be exact, the patients had already presented more or less airway limitation at the time. These tools are not able to play a sufficient role in identifying the healthy subjects at high risk. For instance, as reviewed by Grouse (24), in the study of Bai among Chinese patients, low-dose computed tomography lung scanning diagnosed early COPD when only ~10% of the lung function was affected. Ley-Zaporozhan and Kauczor (25) made an early diagnosis by measuring the airway diameter and wall thickness. Fain *et al* (26) demonstrated presymptomatic detection of degraded pulmonary function in smokers using diffusion-weighted ^{3}He magnetic resonance imaging. These studies have provided information, but a single variable appears to be rather weak to predict the probability of COPD

Table IV. Definition of variables for logistic regression analysis.

Variables	COPD, n	Control, n	P-value
Group			
1=COPD	331		
0=control		351	
Gender[a]			
1=male	298	326	<0.001
0=female	33	25	
Respiratory infection in childhood[a]			
1=yes	49	15	<0.001
0=no	282	336	
Low birth weight[a]			
1=yes	30	2	<0.001
0=no	301	349	
Environmental pollution			
1=yes	103	139	
0=no	228	212	
Smoking[a]			
1=yes	285	214	<0.001
0=no	46	137	
Family history of lung diseases			
1=yes	42	50	
0=no	289	301	
rs2353397			
CT=1 0	140	144	
TT=0 0	70	179	
CC=0 1	121	28	
rs2070600[a]			
GA=1 0	103	134	<0.01
AA=0 0	12	17	
GG=0 1	213	200	
rs10947233[a]			
GT=1 0	112	135	<0.001
TT=0 0	12	26	
GG=0 1	207	190	
rs1800629[a]			
GA=1 0	35	56	<0.001
AA=0 0	0	6	
GG=0 1	296	289	
rs2241712[a]			
AG=1 0	158	170	<0.001
GG=0 0	81	105	
AA=0 1	92	76	
rs1205[a]			
CT=1 0	168	166	<0.01
TT=0 0	89	124	
CC=0 1	70	61	
rs20541			
TC=1 0	150	137	
CC=0 0	138	184	
TT=0 1	39	30	

[a]Significant variables in the final predictive model. COPD, chronic obstructive pulmonary disease.

Table V. Contingency table for Hosmer-Lemeshow test.

	Group=0		Group=1		
Step no.	Observed	Expected	Observed	Expected	Total
1	63	63.037	5	4.963	68
2	54	55.469	14	12.531	68
3	46	47.648	22	20.352	68
4	47	40.928	21	27.072	68
5	37	36.028	31	31.972	68
6	28	31.752	40	36.248	68
7	26	27.886	42	40.114	68
8	24	23.280	44	44.720	68
9	19	17.677	50	51.323	69
10	7	7.296	61	60.704	68

development. A predictive model is required to estimate the risk prior to onset of the disease. The present model possibly aids to calculate the estimation.

Certain previous studies regarding prediction in the fields of COPD may be taken as examples, but they do not refer to the pathogenesis. Schembri *et al* (27) created a model to evaluate the risk of hospitalization and mortality in COPD patients. Castaldi *et al* (28) set up predictive models for FEV1 and the presence of severe COPD in α-1-antitrypsin deficiency, as this information could be used to inform treatment and monitoring decisions. Bacteria play a leading role in acute exacerbations of COPD. A simple prediction model developed by Lode *et al* (29) based on certain factors can identify patients at low risk for exacerbations with gram-negative enteric bacilli and *Pseudomonas aeruginosa*. To the best of our knowledge, a model for COPD development in Chinese patients has not been generated except for the present study.

The present mathematical formula aids in the comprehension of the risk of an individual for whether they smoke or not, as the model includes genetic data summarized from genotyping 76 SNPs in addition to demographic and clinical information. Genetic polymorphisms must be taken into consideration, as COPD is a result of an interaction of genetics and environment. The present case-control study verified that the rs2353397 C allele (*HHIP*), rs1800629 G allele (*TNF-α*), rs2241712 A allele (*TGF-β1*), rs1205 C allele (*CRP*), rs20541 T allele (*IL-13*), rs2070600 G allele (*AGER*) and rs10947233 G allele (*PPT2*) were the risk allelic genes for COPD in a Chinese population. The *HHIP* gene encodes a glycoprotein that is a critical regulator of the hedgehog signaling pathway. The pathway has been indicated in development, repair and cancer in multiple tissues (30). Several gene studies regarding *TNF-α* SNPs also identified that its promoter polymorphism was associated with chronic bronchitis or the extent of emphysematous changes, among which two were carried out in the Caucasian population (31,32) and two in the Japanese population (33,34). The *TGF-β1* SNPs has been explored in the study by Su *et al* (35), which revealed that more COPD patients carried the -800A allele and fewer carried the -509T allele, but there were only 84 COPD and 97 controls who participated in the study. The *IL-13* SNPs, rs2066960, rs20541 and rs1295685, were associated with the COPD risk and a lower

Table VI. Validation of the predictive model.

No.	Group	Gender	Respiratory infection	Low birth weight	Smoking	rs207060	rs10947233	rs10947233	rs1800629	rs2241712	rs1205	Model value
1	1	1	0	0	1	0	0	1	0	0	1	0.57
2	1	1	1	0	1	1	1	0	1	0	1	0.23
3	1	1	0	0	1	0	0	1	0	1	0	0.54
4	1	0	0	0	1	0	0	1	1	0	0	0.23
5	1	0	0	0	0	1	1	0	0	1	0	0.76
6	1	1	0	0	1	1	1	0	0	1	1	0.53
7	1	1	0	0	1	0	0	1	0	0	0	0.66
8	1	1	0	0	1	0	0	1	0	0	0	0.66
9	1	1	0	0	0	0	0	1	0	1	0	0.88
10	1	1	0	0	1	1	1	0	0	1	1	0.53
11	1	1	0	0	1	0	0	1	0	0	0	0.66
12	1	0	0	0	1	0	0	1	0	1	1	0.19
13	1	1	0	0	1	0	0	1	0	0	0	0.66
14	1	1	0	0	1	0	0	1	0	0	1	0.57
15	1	1	0	0	1	0	0	1	0	1	0	0.54
16	1	1	0	0	1	1	1	0	1	0	0	0.59
17	1	1	0	0	1	1	1	0	0	1	0	0.62
18	1	1	0	0	1	0	0	1	0	0	0	0.66
19	1	1	0	0	0	0	0	1	1	0	1	0.81
20	1	1	0	0	1	1	1	0	0	0	1	0.65
21	1	1	0	0	1	0	0	1	0	1	0	0.54
22	1	1	0	0	1	1	1	0	1	0	1	0.50
23	1	0	0	0	1	0	0	1	0	0	0	0.37
24	1	1	0	0	1	1	1	0	1	0	0	0.59
25	1	1	0	0	1	1	1	0	0	1	0	0.62
26	1	1	0	0	1	1	1	0	0	1	0	0.62
27	1	1	0	0	1	1	1	0	0	0	1	0.65
28	1	1	0	0	0	1	1	0	0	0	0	0.95
29	1	1	0	0	0	0	0	1	0	0	0	0.93
30	1	1	0	0	0	0	0	1	0	1	0	0.88
31	0	1	1	0	1	1	1	0	0	1	1	0.25
32	0	1	0	0	1	1	1	0	0	0	1	0.65
33	0	0	1	0	0	0	0	1	0	0	1	0.43
34	0	0	0	0	0	1	1	0	0	1	1	0.68
35	0	0	0	0	1	1	1	0	0	0	0	0.45
36	0	1	0	0	0	0	0	1	1	1	1	0.72
37	0	1	0	0	1	1	1	0	1	1	0	0.46
38	0	0	0	0	0	0	0	1	1	1	1	0.43
39	0	1	1	0	1	0	0	1	0	1	1	0.19
40	0	1	0	0	1	0	0	1	1	1	0	0.38
41	0	1	1	0	1	0	0	1	0	0	0	0.37
42	0	1	0	0	1	0	0	1	1	1	1	0.29
43	0	1	0	1	1	1	1	0	0	0	0	0.20
44	0	1	0	0	1	0	0	1	1	1	0	0.38
45	0	1	0	0	1	0	1	0	1	1	0	0.21
46	0	1	1	0	1	1	1	0	0	1	0	0.34
47	0	1	0	0	1	0	0	1	1	1	0	0.38
48	0	1	1	0	1	1	1	0	1	1	1	0.15
49	0	1	0	0	1	0	0	1	1	1	0	0.38
50	0	1	0	0	1	0	0	1	1	1	1	0.29

baseline lung function in Caucasian patients based on the study by Beghé *et al* (36). The same SNPs as Beghé *et al* were chosen to analyze, but the present results only showed that rs20541 may be of significance in susceptibility in the Chinese population. Sunyer *et al* (37) assessed the association between *CRP* SNP (rs1205) and lung function, and identified

that the TT homozygote in the *CRP* gene was associated with improved lung function. The present results identified that the TT genotype protects patients against COPD, which is similar to the study by Sunyer *et al*, as COPD is characterized by airflow limitation according to lung function. Based on these findings, further research is required to improve the understanding of the gene function in the pathogenesis of COPD. In all the predictive genetic variants that reached the levels of significance in the univariate analysis, five SNPs (rs2070600, rs10947233, rs1800629, rs2241712 and rs1205) were retained through the stepwise variable selection procedure and were incorporated into the final predictive model.

The present study had certain limitations. First, with a larger study sample size, the mathematical formula would have improved the prediction accuracy. Second, further validation in a much larger population is required. Third, although 97 SNPs were selected for the study of genetic susceptibility, further GWAS are required in the Chinese population in order to identify more associated loci, as it is likely that more genetic risk factors would enter the final model.

In conclusion, the present study has established a predictive model for COPD development in a Chinese population, but there remains room for improvement in predictive accuracy. Larger sample sizes for model development and validation will allow for the production of more powerful risk prediction tools.

Acknowledgements

The authors acknowledge the 11th Chinese National Five-Year Development Plan for support of the present study.

References

1. Rabe KF, Beghé B, Luppi F and Fabbri LM: Update in chronic obstructive pulmonary disease 2006. Am J Respir Crit Care Med 175: 1222-1232, 2007.
2. Murray CJL and Lopez AD: Evidence-based health policy - lessons from the Global Burden of Disease Study. Science 274: 740-743, 1996.
3. Celli BR, MacNee W, Agusti A, *et al*; ATS/ERS Task Force: Standards for the diagnosis and treatment of patients with COPD: A summary of the ATS/ERS position paper. Eur Respir J 23: 932-946, 2004.
4. Wang X, Li L, Xiao J, Jin C, Huang K, Kang X, Wu X and Lv F: Association of ADAM33 gene polymorphisms with COPD in a northeastern Chinese population. BMC Med Genet 10: 132-138, 2009.
5. Zhong N, Wang C, Yao W, *et al*: Prevalence of chronic obstructive pulmonary disease in China: A large, population-based survey. Am J Respir Crit Care Med 176: 753-760, 2007.
6. Nihlén U, Montnémery P, Lindholm LH and Löfdahl CG: Detection of chronic obstructive pulmonary disease (COPD) in primary health care: Role of spirometry and respiratory symptoms. Scand J Prim Health Care 17: 232-237, 1999.
7. Lindberg A, Bjerg A, Rönmark E, Larsson LG and Lundbäck B: Prevalence and underdiagnosis of COPD by disease severity and the attributable fraction of smoking Report from the Obstructive Lung Disease in Northern Sweden Studies. Respir Med 100: 264-272, 2006.
8. Tinkelman DG, Price D, Nordyke RJ and Halbert RJ: COPD screening efforts in primary care: What is the yield? Prim Care Respir J 16: 41-48, 2007.
9. Miravitlles M, Soriano JB, García-Río F, Muñoz L, Duran-Tauleria E, Sanchez G, Sobradillo V and Ancochea J: Prevalence of COPD in Spain: Impact of undiagnosed COPD on quality of life and daily life activities. Thorax 64: 863-868, 2009.
10. Shahab L, Jarvis MJ, Britton J and West R: Prevalence, diagnosis and relation to tobacco dependence of chronic obstructive pulmonary disease in a nationally representative population sample. Thorax 61: 1043-1047, 2006.
11. Calverley PM: COPD: Early detection and intervention. Chest 117 (Suppl 2): 365S-371S, 2000.
12. Hersh CP, DeMeo DL, Al-Ansari E *et al*: Predictors of survival in severe, early onset COPD. Chest 126: 1443-1451, 2004.
13. Nizet TA, van den Elshout FJ, Heijdra YF, *et al*: Survival of chronic hypercapnic COPD patients is predicted by smoking habits, comorbidity, and hypoxemia. Chest 127: 1904-1910, 2005.
14. Silverman EK: Progress in chronic obstructive pulmonary disease genetics. Proc Am Thorac Soc 3: 405-408, 2006.
15. Løkke A, Lange P, Scharling H, Fabricius P and Vestbo J: Developing COPD: A 25 year follow up study of the general population. Thorax 61: 935-939, 2006.
16. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, Pukkala E, Skytthe A and Hemminki K: Environmental and heritable factors in the causation of cancer - analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med 343: 78-85, 2000.
17. McCloskey SC, Patel BD, Hinchliffe SJ, Reid ED, Wareham NJ and Lomas DA: Siblings of patients with severe chronic obstructive pulmonary disease have a significant risk of airflow obstruction. Am J Respir Crit Care Med 164: 1419-1424, 2001.
18. Wilk JB, Chen TH, Gottlieb DJ, *et al*: A genome-wide association study of pulmonary function measures in the Framingham Heart Study. PLoS Genet 5: e1000429, 2009.
19. Pillai SG, Ge D, Zhu G, *et al*; ICGN Investigators: A genome-wide association study in chronic obstructive pulmonary disease (COPD): Identification of two major susceptibility loci. PLoS Genet 5: e1000421, 2009.
20. Rabe KF, Hurd S, Anzueto A, *et al*; Global Initiative for Chronic Obstructive Lung Disease: Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. Am J Respir Crit Care Med 176: 532-555, 2007.
21. Gong Y, Shi GC, Wan HY *et al*: Reinvestigation of prevalence of chronic obstructive pulmonary disease in shanghai urban district. J Shanghai Jiaotong Univ (Med Sci) 31: 100-104, 2011.
22. Stratelis G, Jakobsson P, Molstad S and Zetterstrom O: Early detection of COPD in primary care: Screening by invitation of smokers aged 40 to 55 years. Br J Gen Pract 54: 201-206, 2004.
23. Stratelis G, Mölstad S, Jakobsson P and Zetterström O: The impact of repeated spirometry and smoking cessation advice on smokers with mild COPD. Scand J Prim Health Care 24: 133-139, 2006.
24. Grouse L: New studies address urgent need for early COPD diagnosis. J Thorac Dis 4: 19-21, 2012.
25. Ley-Zaporozhan J and Kauczor HU: Imaging of airways: Chronic obstructive pulmonary disease. Radiol Clin North Am 47: 331-342, 2009.
26. Fain SB, Panth SR, Evans MD, Wentland AL, Holmes JH, Korosec FR, O'Brien MJ, Fountaine H and Grist TM: Early emphysematous changes in asymptomatic smokers: Detection with ^{3}He MR imaging. Radiology 239: 875-883, 2006.
27. Schembri S, Anderson W, Morant S, Winter J, Thompson P, Pettitt D, MacDonald TM and Winter JH: A predictive model of hospitalisation and death from chronic obstructive pulmonary disease. Respir Med 103: 1461-1467, 2009.
28. Castaldi PJ, DeMeo DL, Kent DM, *et al*: Development of predictive models for airflow obstruction in alpha-1-antitrypsin deficiency. Am J Epidemiol 170: 1005-1013, 2009.
29. Lode H, Allewelt M, Balk S, De Roux A, Mauch H, Niederman M and Schmidt-Ioanas M: A prediction model for bacterial etiology in acute exacerbations of COPD. Infection 35: 143-149, 2007.
30. Villavicencio EH, Walterhouse DO and Iannaccone PM: The sonic hedgehog-patched-gli pathway in human development and disease. Am J Hum Genet 67: 1047-1054, 2000.
31. Sakao S, Tatsumi K, Igari H, Watanabe R, Shino Y, Shirasawa H and Kuriyama T: Association of tumor necrosis factor-alpha gene promoter polymorphism with low attenuation areas on high-resolution CT in patients with COPD. Chest 122: 416-420, 2002.
32. Keicho N, Emi M, Nakata K, Taguchi Y, Azuma A, Tokunaga K, Ohishi N and Kudoh S: Promoter variation of tumour necrosis factor-alpha gene: Possible high risk for chronic bronchitis but not diffuse panbronchiolitis. Respir Med 93: 752-753, 1999.
33. Stankovic MM, Nestorovic AR, Tomovic AM, Petrovic-Stanojevic ND, Andjelic-Jelic MS, Dopudja-Pantic VB, Nagorni-Obradovic LjM, Mitic-Milikic MM and Radojkovic DP: TNF-alpha-308 promotor polymorphism in patients with chronic obstructive pulmonary disease and lung cancer. Neoplasma 56: 348-352, 2009.

34. Papatheodorou A, Latsi P, Vrettou C, *et al*: Development of a novel microarray methodology for the study of SNPs in the promoter region of the TNF-alpha gene: Their association with obstructive pulmonary disease in Greek patients. Clin Biochem 40: 843-850, 2007.
35. Su ZG, Wen FQ, Feng YL, Xiao M and Wu XL: Transforming growth factor-beta 1 gene polymorphisms associated with chronic obstructive pulmonary disease in Chinese population. Acta Pharmacol Sin 26: 714-720, 2005.
36. Beghé B, Hall IP, Parker SG, Moffatt MF, Wardlaw A, Connolly MJ, Fabbri LM, Ruse C and Sayers I: Polymorphisms in IL13 pathway genes in asthma and chronic obstructive pulmonary disease. Allergy 65: 474-481, 2010.
37. Sunyer J, Pistelli R, Plana E, Andreani M, Baldari F, Kolz M, Koenig W, Pekkanen J, Peters A and Forastiere F: Systemic inflammation, genetic susceptibility and lung function. Eur Respir J 32: 92-97, 2008.