A gene-based risk score for lung cancer susceptibility in smokers and ex-smokers

R P Young,[1] R J Hopkins,[1] B A Hay,[1] M J Epton,[2] G D Mills,[3] P N Black,[1] H D Gardner,[1] R Sullivan,[4] G D Gamble[1]

[1] Department of Medicine, Auckland Hospital, Auckland, New Zealand; [2] Department of Medicine, University of Otago, Christchurch, New Zealand; [3] Department of Medicine, Waikato Hospital, Hamilton, New Zealand; [4] Department of Oncology, Auckland Hospital, Auckland, New Zealand

Correspondence to:
Dr R Young, Department of Medicine, Auckland Hospital, Private Bag 92019, Auckland, New Zealand; roberty@adhb.govt.nz

ABSTRACT

Background: Epidemiological and family studies suggest that lung cancer results from the combined effects of age, smoking and genetic factors. Chronic obstructive pulmonary disease (COPD) is also an independent risk factor for lung cancer and coexists in 40–60% of lung cancer cases.

Methods: In a two-stage case–control association study, genetic markers associated with either susceptibility or protection against lung cancer were identified. In a test cohort of 439 Caucasian smokers or ex-smokers, consisting of healthy smokers and lung cancer cases, 157 candidate single nucleotide polymorphisms (SNPs) were screened. From this, 30 SNPs were identified, the genotypes (codominant or recessive model) of which were associated with either the healthy smokers (protective) or lung cancer (susceptibility) phenotype. After genotyping of this 30-SNP panel in a second validation cohort of 491 subjects and using the same protective and susceptibility genotypes from our test cohort, a 20-SNP panel was selected on the basis of independent univariate analyses.

Results: Using multivariate logistic regression, including the 20 SNPs, it was also found that age, history of COPD, family history of lung cancer and gender were significantly and independently associated with lung cancer.

Conclusions: When numeric scores were assigned to both the SNP and demographic data, and sequentially combined by a simple algorithm in a risk model, the composite score was found to be linearly related to lung cancer risk with a bimodal distribution. Genetic data may therefore be combined with other risk variables from smokers or ex-smokers to identify individuals who are most susceptible to developing lung cancer.

Approximately 90% of people with lung cancer have a smoking history, yet only 10–15% of chronic smokers get lung cancer, suggesting that factors in addition to smoking exposure must be relevant.[1] Epidemiological studies have identified age, smoking exposure, impaired lung function and family history as key risk factors for lung cancer.[2] Genetic factors have been shown to play a modest role in determining susceptibility to lung cancer,[3 4] most likely by conferring an inherent susceptibility (exaggerated or maladaptive response) to chronic inflammation from aero-pollutant exposure (most commonly smoking).[5 6] Like many cancers, this provides the initial stimulus to tissue remodelling (eg, small airway disease and/or emphysema), DNA damage and impaired cell cycle control.[6 7]

Prospective studies have shown that ~20% of smokers develop chronic obstructive pulmonary disease (COPD), defined by spirometry (forced expiratory volume in 1 s (FEV_1)/forced vital capacity <70% and/or FEV_1<80% of predicted), whereas the majority have preserved lung function at, or close to, predicted values.[8] Both prospective and retrospective studies show that spirometric evidence of COPD is found in 40–60% of smokers diagnosed with lung cancer.[9 10] In contrast with smokers who maintain normal lung function, those with COPD have a 2–6-fold greater risk of lung cancer, independent of age and pack-years.[9–11] These studies suggest that smokers with COPD are at an inherently increased risk of lung cancer (susceptible phenotype), whereas most smokers with normal lung function (estimated to be 80%) are at least risk (resistant phenotype).[9–11]

Genetic predisposition to lung cancer is likely to be both polygenic and heterogeneous, conferred by a variable combination of relatively common polymorphisms with low penetrance and modest effect sizes.[12 13] Moreover, it is likely that important smoking-gene interactions underlie lung cancer,[14] as seen in other smoking-related cancers (eg, bladder and stomach). As genetic variants associated with both COPD and lung cancer have been identified, and include the recently reported chromosome 15q25 gene locus,[15 16] we suggest it is important to measure lung function in all participants of case–control studies of lung cancer. For both clinical and biostatistical reasons, screening the exposed controls will increase the power of the study to identify relevant genetic variants compared with studies in which the control group is unscreened.[17]

It is well known that non-genetic risk factors such as age, history of COPD and smoking history are very important and can be combined to develop risk-based tools for determining lung cancer susceptibility.[18 19] Recently, genotype data from previously implicated prostate cancer susceptibility single-nucleotide polymorphisms (SNPs) were combined with family history to derive risk estimates for prostate cancer.[20] The objective of this study was to use a similar approach to analysing data from a case–control study and show how genetic variants, previously showing small effects on lung cancer risk, can be combined in an algorithm with other known risk factors to derive a gene-based susceptibility score for lung cancer.

METHODS

Study population

This study was a two-stage case–control design conducted in three centres following the same recruitment protocol. Only people of Caucasian

ancestry were recruited (all four grandparents of Caucasian descent). Lung cancer cases were identified through hospital clinics between 2004 and 2007 using the following criteria: >40 years of age, minimum 15 pack-years of smoking, diagnosis confirmed on histological or cytological grounds, and limited to the following four histological subtypes: adenocarcinoma, squamous cell cancer, small cell cancer and non-small cell cancer (generally large cell or bronchoalveolar subtypes). The median time interval between diagnosis and recruitment was 3 months. Patients with lung cancer underwent blood sampling for DNA extraction, an investigator-administered questionnaire and spirometry with a portable spirometer (Easy-One; ndd Medizintechnik AG, Zurich, Switzerland) following American Thoracic Society (ATS) criteria. For patients with lung cancer who had already undergone surgery, results of preoperative lung function tests performed by the hospital laboratory (using ATS criteria) were sourced from the medical records.

Control subjects were recruited from the same communities as the cases using the following criteria: Caucasian ancestry (as defined above), age 45–80 years and past or current smoking history of a minimum of 15 pack-years. Controls were volunteers who met the above criteria and were identified through either a community mail-out or while attending a community-based club for older people located in the suburbs that were the referral base for the hospital clinics from which the lung cancer cases were recruited. All smoking controls underwent blood sampling, spirometry and the same investigator-administered questionnaire given to the cases. Controls with spirometry consistent with COPD were excluded (30% of those who volunteered). Informed written consent was obtained from both lung cancer patients and community-acquired healthy "smoker" controls. The study was approved by the local ethics committee. The questionnaire (modified from the ATS respiratory questionnaire) included data on demographic variables such as age, gender, medical history, family history of lung disease, active and passive tobacco exposure, and occupational aero-pollutant exposures.

Selection and genotyping of SNPs

After extensive review of both the lung cancer and COPD literature, polymorphisms with the following attributes were selected for initial screening in the test cohort: (*a*) SNPs in genes encoding proteins in pathways of cell-cycle control, oxidant response, apoptosis and airways inflammation; (*b*) SNPs known either to have functional effects on in vitro assays or to be either non-synonymous or in regulatory regions. In a test cohort of 439 smokers (run 1 recruited during 2003–2005: 239 lung cancer cases and 200 control smokers), 157 candidate SNPs were screened (available on request), and those where the difference in genotype frequencies exceeded a 20% magnitude difference and p value <0.20 were identified as part of our model-forming approach.[21] Where the call rate (percentage of samples for which genotyping failed for technical reasons) fell below 95% for any cohort, the reading and/or genotyping of failed samples was repeated for that SNP; after retesting, SNPs with call rates <95% were not included in further analysis. SNPs were assigned as "protective" when the homozygote and/or heterozygote genotype for either allele was found more often in control smokers than lung cancer cases (in a recessive or codominant model). SNPs were assigned as "susceptible" when the homozygote and/or heterozygote genotype was found significantly more often in lung cancer cases than control smokers.

Genotyping

Genomic DNA was extracted from whole blood samples using standard salt-based methods. Purified genomic DNA was aliquoted (10 ng/µl concentration) into 96-well plates and genotyped on a Sequenom system (Sequenom Autoflex Mass Spectrometer and Samsung 24 pin nanodispenser) by the Australian Genome Research Facility (www.agrf.com.au) using sequences designed in-house (available on request) and recommended amplification and separation methods (iPLEX; www.sequenom.com).[16]

Of the 157 candidate SNPs screened in our discovery cohort, 30 met the above criteria. These SNPs were genotyped in a second cohort of 491 smokers using identical recruitment methods (run 2 recruited during 2006–2007: 207 lung cancer cases and 284 control smokers). For all SNP assays, again a minimum call rate of 95% was required. This second validation cohort of lung cancer cases and control smokers was identical with the first groups with respect to demographic factors and lung cancer characteristics. On the basis of independent univariate analyses in run 1 and run 2 (consistency, direction and significance of association), a final panel of the 20 most discriminatory SNPs was selected (12 susceptibility SNPs and eight protective SNPs from the test panel of 30).

Algorithm

The assignment of a protective or susceptible SNP genotype was made from the test cohort data (run 1) and was strictly applied to the data from run 2. On the basis of an algorithm derived from our work on the genetics of COPD (unpublished data), a scoring system was applied to the genotypes for each of the susceptibility and protective SNPs. For each subject, a numerical value of −1 was assigned for each of the protective genotypes present among the protective SNPs and +1 for each of the susceptible genotypes present. Where an individual did not have either the protective or susceptibility genotype for that SNP, the score was 0 (ie, did not contribute to the genetic score). This approach is consistent with a recently published study in prostate cancer.[20] Weighting the presence of specific susceptible or protective genotypes according to their individual odds ratios (ORs; from univariate regression) did not significantly improve the discriminatory performance of the raw SNP score (unpublished data).

Lung cancer susceptibility score

The approach of deriving an overall "susceptibility score" by combining independent risk factors is comparable to existing risk-scoring systems such as the Prostate Cancer Test, Framingham Score for coronary artery disease risk and the Gail Score for breast cancer.[18–20 22 23] By using multivariate logistic and stepwise regression analysis, the 20-SNP panel was examined in combination with relevant non-genetic factors. This analysis of run 1 data identified age, family history of lung cancer and previous diagnosis of COPD as significant contributors to lung cancer susceptibility. In addition, and consistent with other case–control studies, female gender in our study was also associated with a small increased risk of lung cancer (p<0.01). However, we did not include gender in the final risk model, as its importance in prospective studies has been lacking.[24] On the basis of a multivariate logistic regression analysis in run 1 (see results for combined analysis below), a score was assigned according to age, history of COPD and family history. These variables have been identified in other risk assessment tools for lung cancer susceptibility[18 19] and improved

the discriminatory power of the SNP score data alone. As smoking exposure (pack-years) was a recruitment criterion for this study and comparable between cases and controls, it was not surprising to find that it made little contribution to this scoring system derived from our cohorts. The lung cancer susceptibility score was plotted with (*a*) the frequency of lung cancer and (*b*) the floating absolute risk (equivalent to OR) across the combined smoker/ex-smoker cohort.[25 26]

Statistical analysis

Patient characteristics in the cases and controls were compared by unpaired t tests for continuous variables and χ^2 test for discrete variables. Genotype and allele frequencies were checked for each SNP by Hardy–Weinberg equilibrium (tests that genotype frequencies were as expected from the allele frequencies). Population admixture was excluded by the population structure analysis on genotyping data from 40 unrelated SNPs.[27] Distortions in the genotype frequencies were identified between cases and controls using 2×3 contingency tables. Genotype data (20-SNP panel) and the most relevant non-genetic variables were combined in a stepwise fashion to assess their combined effects on discriminating low and high risk (by OR and receiver operating characteristic (ROC) curve) by score quintile. The frequency distribution of the optimised lung cancer susceptibility score was compared across the cases and controls. Its clinical utility was assessed using ROC analysis, which assesses how well the model predicts risk across the score (ie, clinical performance of the score with respect to sensitivity, specificity and false positive rate). To assess the stability of the optimised risk model, a sensitivity analysis was performed in which age, gender and smoking dose were more stringently matched between cases and controls. The effect on sensitivity and performance of the lung cancer susceptibility score to the addition of non-genetic variables was also assessed by comparing ORs and ROC analyses.

RESULTS

Demographic variables and genotyping

Table 1 summarises the characteristics of the lung cancer cases and healthy control smokers. The 446 lung cancer cases from run 1 (n = 239) and run 2 (n = 207) were comparable (with respect to demographic characteristics, histology and staging) and similar to a large published series.[28] Given the small difference in age, the 482 healthy control smokers (200 in run 1, 282 in run 2) were comparably exposed with respect to smoking and other aero-pollutants. The lower frequency of current smokers in the lung cancer group probably reflects coexisting COPD (higher quit rates), and longer duration of smoking in lung cancer cases reflects an older age. In a gene-by-smoking interaction model such as this, differences in smoking exposure are more likely to obscure effects (bias to the null) than generate effects. Consistent with the findings of others, the lung cancer cohort had higher rates of a family history of lung cancer (19% vs 9%) and history of COPD (29% vs 5%). The latter (5%) probably reflects a clinical diagnosis of COPD, based on symptoms but not spirometry, in smokers with asthma and/or chronic bronchitis. As expected, lung function was worse in the lung cancer cohort than the healthy smoker controls. Testing lung function in the lung cancer cases (performed within 3 months of diagnosis, in the absence of pleural effusions and before surgery) allows us to test for confounding by COPD (see below).

Table 1 Summary of characteristics for the lung cancer and resistant smokers

Variable	Lung cancer patients (n = 446)	Healthy control smokers (n = 484)	p Value
Characteristic			
Male	53%	60%	0.007
Age (years)	69 (10)	60 (10)	<0.001
Height (cm)	167 (0.08)	170 (0.09)	<0.001
Weight (kg)	69 (15)	79 (15)	<0.001
History of COPD	29%	5%	<0.001
Smoking history			
Non-smokers (%)	7%	0	
Current smoking (%)	35%	48%	<0.001
Age started (years)	18 (4)	17 (3)	0.18
Years smoked	41 (12)	35 (11)	<0.001
Pack-years	41 (25)	40 (19)	0.28
Cigarettes/day	20 (10)	24 (11)	<0.001
History of other exposures			
In utero smoke exposure	18%	17%	0.45
Mother smoked in childhood	37%	41%	0.03
Home ETS exposure as adult	79%	58%	<0.001
Work ETS exposure	86%	63%	<0.001
Work dust exposure	63%	47%	<0.001
Work fume exposure	41%	38%	0.16
Asbestos exposure	23%	16%	0.02
Family history			
COPD	33%	28%	0.12
Lung cancer	19%	9%	<0.001
Lung function			
FEV$_1$ (litres)	1.86 (0.48)	2.86 (0.68)	<0.001
FEV$_1$ % predicted	73%	99%	<0.001
FEV$_1$/FVC	64 (13)	78 (7)	<0.001
Spirometric COPD*	51%	0%	<0.001

Values are frequency (given as %) or mean (SD).
*According to GOLD 2+ criteria (FEV$_1$/FVC < 70% and FEV$_1$ % predicted ⩽ 80%).
COPD, chronic obstructive pulmonary disease; ETS, environmental tobacco smoke; FEV$_1$, forced expiratory volume in 1 s; FVC, forced vital capacity.

The observed genotypes for the 20 SNPs in this study were in Hardy–Weinberg equilibrium (table 2), thereby excluding significant genotyping error. The genotype frequencies for the controls were comparable to those from the International Hapmap Project (www.hapmap.org). The development of the lung cancer susceptibility score is described in the Methods section, and a summary of the 20-SNP panel univariate analysis is presented in table 3. Although six of the top 20 SNPs do not reach traditional levels of significance, they have been included in the panel because (*a*) in previous studies they have been shown to have functional effects, (*b*) they have been associated with COPD and/or lung cancer (see Discussion), (*c*) in combination they make a contribution to the performance of the susceptibility score, and (*d*) their inclusion recognises the likely genetic heterogeneity that exists in lung cancer case–controls studies. A SAS macro was used to estimate the false discovery rate (FDR) (Osborne JA, North Carolina State University; http://www2.sas.com/proceedings/sugi31/190-31.pdf) and produce a q statistic as the smallest p value that would be said to be statistically significant while preserving an overall 5% significance level.

Table 2 Expected genotype frequencies and Hardy–Weinberg equilibrium (HWE)

| SNP No | SNP name (rs) | Allele frequencies (study total = 930) | | HWE observed genotypes (p Value) | Allele frequencies | | | |
| | | | | | Study controls (n = 484) | | Hapmap Caucasian* | |
		Major	Minor		Major	Minor	Major	Minor
1	A5 nAChR (rs 16969968)	0.65	0.35	0.57	0.69	0.31	0.58	0.42
2	CYP 2E1 (rs 2031920)	0.98	0.02	0.06	0.99	0.01	0.94	0.06
3	Interleukin-18 (rs 360721)	0.70	0.30	0.77	0.68	0.32	0.70	0.32
4	Interleukin-8 (rs 4073)	0.54	0.46	0.14	0.51	0.49	0.60	0.40
5	Interleukin 1B (rs 16944)	0.69	0.31	0.80	0.67	0.33	0.65	0.35
6	ITGA11 (rs 2306022)	0.90	0.10	0.07	0.91	0.09	0.93	0.07
7	NAT 2 (rs 1799930)	0.71	0.29	0.87	0.69	0.31	0.71	0.29
8	α1-antichymotrypsin (rs 4934)	0.50	0.50	0.54	0.47	0.53	0.51	0.49
9	Cerberus 1 (rs 10115703)	0.92	0.08	0.54	0.93	0.07	0.89	0.11
10	DAT1 (rs 6413429)	0.93	0.07	0.78	0.94	0.06	0.87	0.13
11	TNFR1 (rs 1139417)	0.57	0.43	0.96	0.56	0.44	0.51	0.49
12	TLR9 (rs 5743836)	0.85	0.15	0.92	0.85	0.15	0.84	0.16
13	P73 (TP73) (rs 2273953)	0.75	0.25	0.93	0.78	0.22	0.85	0.15
14	SOD3 (rs 1799895)	0.99	0.01	0.99	0.98	0.02	0.97	0.03
15	ITGB3 (rs 2317676)	0.93	0.07	0.51	0.91	0.09	0.95	0.05
16	DRD2 (rs 1799732)	0.90	0.10	0.67	0.89	0.11	0.90	0.10
17	BCL2 (rs 2279115)	0.51	0.49	0.60	0.53	0.47	0.57	0.43
18	XPD (ERCC2) (rs 13181)	0.61	0.39	0.90	0.59	0.41	0.67	0.33
19	REV1 (REV1L) (rs 3087386)	0.56	0.44	0.96	0.58	0.42	0.50	0.50
20	FasL (TNFSF6) (rs 763110)	0.63	0.37	0.83	0.61	0.39	0.64	0.36

*Allele frequencies for Caucasians from www.hapmap.org.
SNP, single nucleotide polymorphism.

Risk model development

In a multivariate logistic regression analysis that included the SNPs (individually), age (>60 years), family history of lung cancer (first-degree relative), gender and history of COPD, the OR for the susceptibility and protective SNPs was 1.1–3.2 and 0.20–0.80, respectively (the combined SNP score is independently related to lung cancer, p<0.001). The OR for age >60 years, family history of lung cancer and history of COPD was 3.5 (95% CI 2.5 to 4.9, p<0.001), 2.5 (95% CI 1.6 to 4.0, p<0.001) and 7.5 (95% CI 4.5 to 12.4, p<0.001), respectively (total area under the curve (AUC) = 0.80 where SNPs were included individually with adjustment for all variables). History of COPD in this model confers a high risk, in part due to differences in lung function derived from the study design. On the basis of these findings, and those from previously published studies,[3–4] [9–11] we derived an optimised score by assigning scores to non-genetic variables as follows; +4 for those aged >60 years old, +3 for those with a family history of lung cancer and +4 for those with a diagnosis of COPD (ie, age and diagnosis of COPD equally weighted).

Lung cancer score = (number of susceptible genotypes) − (number of protective genotypes) + 3 (for positive family history for lung cancer) + 4 (for past diagnosis of COPD) + 4 (for age>60 years old).

Such an approach is consistent with existing risk scores[18] [19] and places the SNP data in an appropriate clinical context.[22] [23] Gender was not included in the finalised risk model for the reasons described above (and its inclusion did not alter the AUC).

Model performance

In the optimised model, the lung cancer susceptibility score was compared with frequency of lung cancer, and a linear relationship was found across the lung cancer susceptibility scores ≤1 to 8+, with lung cancer frequency spanning 17–86% (fig 1). The magnitude of this effect was also sequentially examined using the floating absolute risk[25] [26] plotted on a log scale (equivalent to

an OR), which references the lowest frequency group as OR = 1 (referent group) and compares the lung cancer score with the referent group. The OR for SNPs alone (fig 2a), SNPs, family history and age (fig 2b) and SNPs, family history, age and COPD (fig 2c) spanned from 1 to 10 (p<0.001), 1 to 19 (p<0.001) and 1 to 28 (p<0.001), respectively, across the lung cancer scores when subjects were grouped approximately as heptiles or quintiles. Subgrouping by age band or histology did not alter this linear relationship between score and OR (data not shown).The lung cancer susceptibility score for lung cancer cases and controls shows a bimodal distribution on frequency distribution, indicating potential utility as screening test.[29]

Analysis of model sensitivity

To correct for the small differences in age, smoking status, COPD and gender mix between cases and controls, a subgroup (sensitivity) analysis was performed (a) limited to those >60 years of age (age weighting equally applied to all), (b) removing COPD from the model, and (c) where mean age, pack-years and gender were closely matched between cases and controls (n = 450: 72 vs 69 years, 45 vs 43 pack-years and 70% vs 70% male respectively). A linear increase in OR across quintiles of the lung cancer susceptibility score (range 1–58, p<0.01) was still evident, with confidence intervals consistent (ie, overlapping) with those derived with the full dataset (fig 2).

ROC analysis

In a ROC analysis (n = 930) of the optimised model, we found that the AUC or c statistic for run 1, run 2 and run 1+2 was 0.82, 0.75 and 0.79, respectively. The AUC in the total cohort for the 20-SNP panel, age, family history of lung cancer and history of COPD on their own were 0.68, 0.70, 0.55 and 0.62, respectively. When just "genetic factors" are used in the risk model (SNPs + family history of lung cancer), as seen in the Prostate Cancer Study,[20] the ORs span 1–10 across septiles and the AUC = 0.70 (with no contribution from age and COPD).

Table 3 Genotypes and results of univariate analysis

SNP (rs No)*	Genotype	Lung cancer	Smoking controls	Call rate (%)	Univariate OR (95% CI)	p Value	Phenotype
α5-nAChR (rs 16969968)	AA	68 (16%)	45 (9%)		1.8 (1.2 to 2.7)	0.004	Susceptibility
	AG/GG	361 (84%)	426 (91%)				
CYP 2E1 (rs 2031920)	TT/TC	24 (6%)	14 (3%)	95	2.1 (1.0 to 4.3)	0.03	Susceptibility
	CC	379 (94%)	463 (97%)				
Interleukin-18 (rs 360721)	CC	237 (54%)	208 (45%)	96	1.4 (1.1 to 1.9)	0.009	Susceptibility
	CG/GG	201 (46%)	250 (55%)				
Interleukin-8 (rs 4073)	TT	129 (31%)	109 (23%)	96	1.5 (1.1 to 2.1)	0.005	Susceptibility
	AT/AA	284 (69%)	367 (77%)				
Interleukin 1B (rs 16944)	GG	215 (49%)	212 (44%)	99	1.2 (0.9 to 1.6)	0.14	Susceptibility
	AA/AG	224 (51%)	269 (56%)				
ITGA11 (rs 2306022)	AA	14 (3%)	6 (1%)	98	2.6 (0.9 to 7.6)	0.04	Susceptibility
	GA/GG	422 (97%)	470 (99%)				
N-Acetylcysteine transferase 2 (rs 1799930)	GG	239 (56%)	222 (47%)	97	1.4 (1.1 to 1.9)	0.006	Susceptibility
	AA/AG	189 (44%)	253 (53%)				
α1-Antichymotrypsin (rs 4934)	GG	123 (28%)	96 (20%)	98	1.6 (1.2 to 2.2)	0.004	Susceptibility
	AG/AA	312 (72%)	383 (80%)				
Cerberus 1 (rs 10115703)	AA/AG	71 (16%)	59 (12%)	97	1.4 (0.9 to 2.0)	0.10	Susceptibility
	GG	363 (84%)	413 (88%)				
DAT1 (rs 6413429)	GT/TT	64 (15%)	50 (10%)	98	1.5 (1.0 to 2.3)	0.04	Susceptibility
	GG	367 (85%)	431 (90%)				
TNFR1 (TNFRSF1A) (rs 1139417)	AA	148 (36%)	142 (30%)	96	1.3 (1.0 to 1.8)	0.05	Susceptibility
	AG/GG	258 (64%)	329 (70%)				
TLR9 (rs 5743836)	CC	12 (3%)	6 (1%)	96	2.2 (0.8 to 6.6)	0.12	Susceptibility
	CT/TT	419 (97%)	455 (99%)				
P73 (TP73) (rs 2273953)	CC	219 (52%)	292 (62%)	96	0.65 (0.49 to 0.85)	0.001	Protective
	TC/TT	206 (48%)	178 (38%)				
SOD3 (rs 1799895)	GG/GC	4 (1%)	15 (3%)	96	0.28 (0.10 to 0.90)	0.02	Protective
	CC	425 (99%)	451 (97%)				
ITGB3 (rs 2317676)	GG/GA	44 (10%)	77 (16%)	98	0.59 (0.39 to 0.89)	0.008	Protective
	AA	391 (90%)	403 (84%)				
DRD2 (rs 1799732)	CDel/Del.Del	70 (16%)	107 (22%)	98	0.68 (0.48 to 0.96)	0.02	Protective
	CC	359 (84%)	372 (78%)				
BCL2 (rs 2279115)	AA	103 (24%)	145 (31%)	97	0.71 (0.53 to 0.97)	0.03	Protective
	AC/CC	328 (76%)	330 (69%)				
XPD (ERCC2) (rs 13181)	GG	60 (14%)	81 (18%)	96	0.74 (0.51 to 1.10)	0.11	Protective
	GT/TT	376 (86%)	377 (82%)				
REV1 (REV1L) (rs 3087386)	CC	128 (29%)	163 (34%)	98	0.79 (0.59 to 1.10)	0.10	Protective
	TC/TT	310 (71%)	312 (66%)				
FasL (TNFSF6) (rs 763110)	TT	53 (12%)	78 (16%)	98	0.72 (0.49 to 1.10)	0.09	Protective
	TC/CC	379 (88%)	403 (84%)				

*OMIM nomenclature.
SNP, single nucleotide polymorphism.

On stepwise analysis, age and the SNP panel make the greatest contribution to the AUC (SNPs = 0.68, age + SNPs = 0.76, age + SNPs + family history = 0.77), with history of COPD making a small additional contribution (total AUC = 0.79) (fig 3). Using an FDR analysis, 12 SNPs were identified as being significantly associated with lung cancer and, when combined with age and family history, derived an AUC = 0.75. When gender was included in the model, the AUC was not improved.

Inclusion of COPD

As smokers with normal lung function were selected as controls (resistant or lowest risk phenotype), and history of COPD was included in the model, it is necessary to examine the effect of including COPD in the model. In the ROC analysis, history of COPD alone was a modest discriminator (AUC = 0.62) and added little to the other variables in the combined model (increased AUC from 0.77 to 0.79). When history of COPD was removed from the model, the ORs span 1–19 across quintiles (p<0.001) and performance characteristics are minimally affected (AUC = 0.77). The model was also tested in young smokers (≤60 years old), in whom COPD prevalence was only 3% and the ORs spanned 1–16 (p<0.001). Most importantly, the model was assessed by comparing the smoking controls and lung cancer cases subgrouped according to lung function (fig 4a,b). This shows that (a) the distribution of the susceptibility score was comparable between lung cancer cases divided by those with high or low lung function (fig 4a) and (b) when people with COPD (based on spirometry) are removed from the analysis (leaving smoking controls compared with lung cancer cases with normal lung function), the bimodal distribution is not affected (fig 4b). We conclude that the risk model is not significantly affected after adjustment for differences in COPD prevalence between cases and controls.
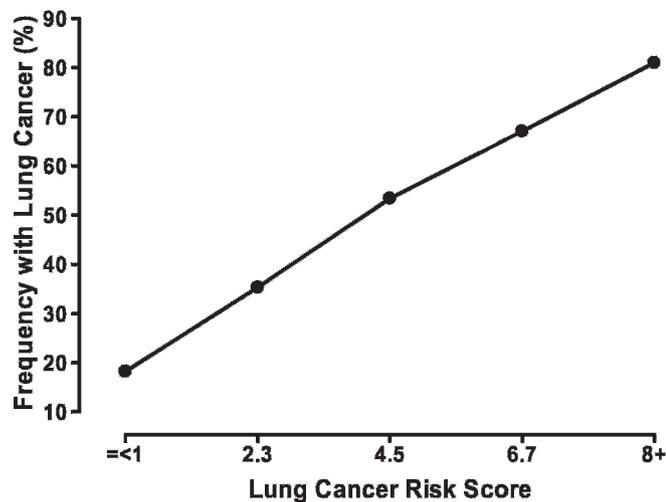
Figure 1 Frequency of lung cancer according to the lung cancer susceptibility (risk) score modelled with single-nucleotide polymorphisms, age, family history and chronic obstructive pulmonary disease.

DISCUSSION

This study has used a two-stage case–control candidate gene approach and identified a panel of protective and susceptibility SNPs that individually confer only small effects (OR ranging from 0.3 to 2.6). This is very much in keeping with the experience from case–control association studies to date.[30–62] Consistent with existing risk models, relevant factors were combined using an algorithm (in this study including SNP data) to derive a susceptibility score on a simple linear scale. This study design, and the algorithmic approach that underlies our lung cancer susceptibility score, takes into account important epidemiological observations relevant to genetic predisposition to lung cancer. Firstly, that, although smoking exposure is for the majority a prerequisite to developing lung cancer, increasing age, smoking dose and poor lung function have important independent effects on lung cancer susceptibility. Secondly, the genetic factors underlying lung cancer risk are likely to be both polygenic and heterogeneous, conferred by a variable combination of genetic variants (ie, SNPs with low penetrance and small effect sizes). Thirdly, genetic factors may confer either a protective[30 31] or susceptibility[16] phenotype to lung cancer. Here we report a 20-SNP panel which, combined with family history,[20] define risk (OR) across quintiles ranging from 1 to 10 with an AUC of 0.70. A risk tool with greater clinical utility can be derived by including age and presence of COPD to identify those at greatest susceptibility to lung cancer (OR range 1–28 and AUC = 0.79).

Several other important factors relevant to the genetic epidemiology of lung cancer have been considered in the design of this study. We sought to minimise false-positive results in a number of ways. The most important of these was to internally validate our findings using a two-stage design with an initial test cohort (run 1) to identify SNPs of potential interest. We then tested only those SNPs in a second cohort of cases and controls (run 2) using univariate and multivariate analysis to rank the SNPs under both conditions. Secondly, population stratification was excluded, and, thirdly, the presence of genotyping error was minimised through Hardy–Weinberg equilibrium analysis (see Methods) and by the exclusion of SNPs with <95% call rate (fails on genotyping are invariably
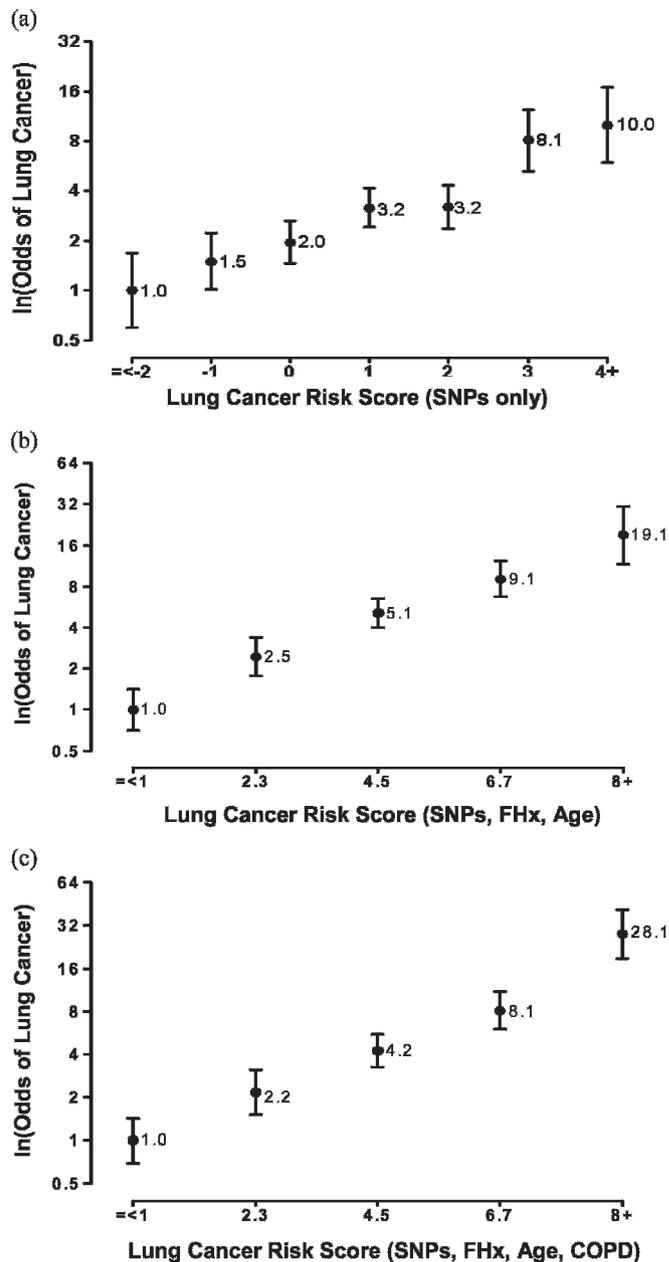

Figure 2 Odds ratio of lung cancer according to the lung cancer susceptibility (risk) score using (a) single-nucleotide polymorphisms (SNPs) only, (b) SNPs, family history (FHx) and age, and (c) SNPs, family history, age and history of chronic obstructive pulmonary disease (COPD).

genotype specific, thus generating false-positive associations). With respect to important confounding factors, our lung cancer cases and healthy smoking controls were matched for smoking exposure (pack-years). They were also similar with respect to gender and age mix. When the combined cohort was sub-grouped by age band, the lung cancer susceptibility score maintained its discriminating utility across all groups. It was concluded that lung function was not confounding the results of this study, as the distribution of the lung cancer susceptibility score across the lung cancer cases, subdivided by normal and low lung function, showed no significant difference. The same could not be said of previously published case–control studies in which lung function was not measured.

However, weaknesses in this study include the modest size of the cohorts, borderline significance of some SNPs in the absence

Figure 3 Distribution of the lung cancer susceptibility (risk) score between cases and controls. (a) Single-nucleotide polymorphisms (SNPs) only, (b) SNPs, family history and age, and (c) SNPs, family history, age and history of chronic obstructive pulmonary disease. Ca, cancer.
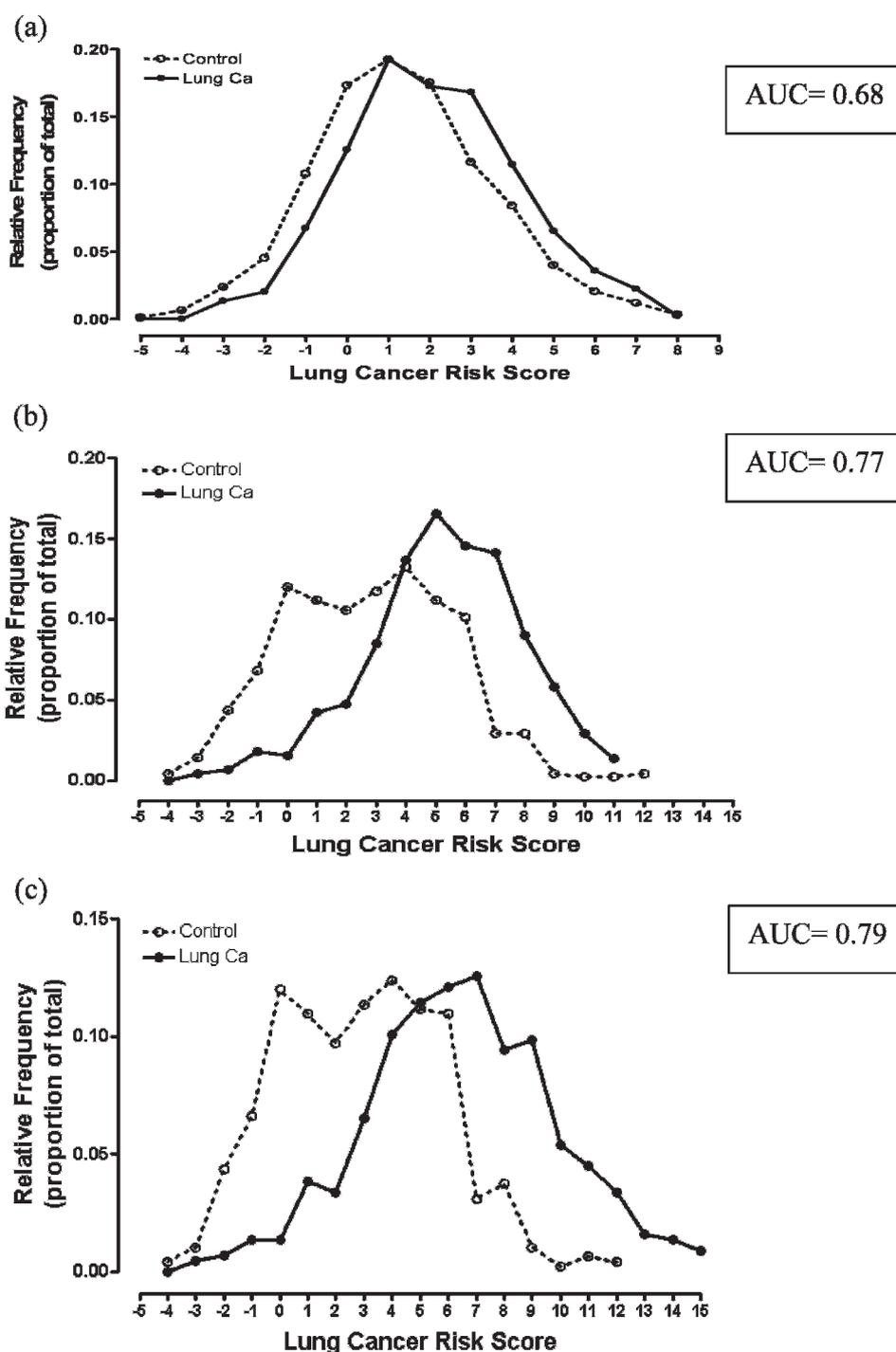
of correction, cross-sectional design, and recruitment limited to Caucasians. Moreover, it is accepted that, by selecting a control population with normal lung function (but comparable exposure) and including COPD (history) in the score, we will increase the score in those with lung cancer compared with controls (5% vs 29%). However, although this increases the magnitude of the difference between cases and controls (reflected in the ORs), it contributes little to the performance of the score (adds 0.02 to the AUC; see the Results section). Moreover, when subjects with COPD are excluded from the analysis (fig 4b, smoking controls versus lung cancer cases with normal lung function), the discriminating utility of the score is unaffected. In addition, in the youngest age band (confined to cases and controls $\leqslant 60$ years old), the prevalence of COPD (history) was only 3% (little effect from COPD weighting), there was no age weighting, and the susceptibility score was still a good discriminator (OR spans 1–16, p<0.001). We argue that screening individuals with COPD (based on spirometry) out of the controls has the following advantages: (1) best reflects the majority of smokers with no COPD estimated at 80%[8]; (2) best reflects the majority of smokers who will not develop lung cancer (resistant phenotype) estimated to be 80–90% (thereby minimising the dilutional effects of including patients with COPD, ie, misclassification)[1 17]; (3) best suited to identifying "protective" SNPs by comparing exposed individuals at either end of the risk spectrum.[30 31] That said, replication
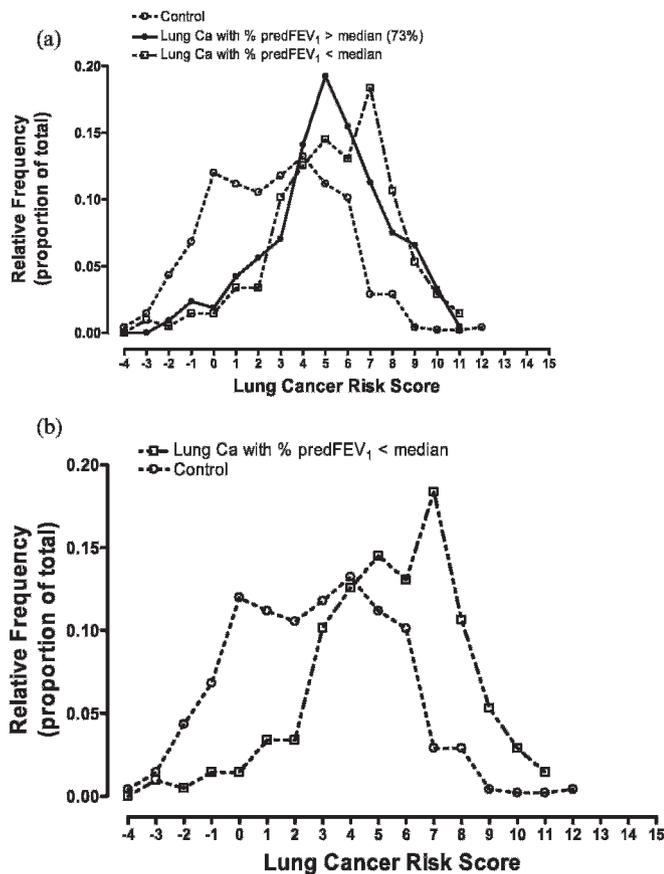
Figure 4 (a) Frequency distribution of the lung cancer score among controls and lung cancer cases divided according to low and normal lung function. (b) Frequency distribution of lung cancer score among controls and lung cancer cases with normal lung function. Ca, cancer; FEV_1, forced expiratory volume in 1s.

using an unselected control group (in which COPD prevalence would be 10% or more) might better reflect an unselected at-risk population and, as expected, reduce differences in the susceptibility score between cases and controls (dilutional effect). Although population stratification was formally tested, and our population is confined to Caucasians (where population admixture is less of an issue), it is possible that this remains a problem. A further limitation of the study is that, although the cases and controls were arguably representative, not all variables were precisely matched (eg, age, gender and smoking patterns). We reanalysed our data in a closely matched cohort (n = 450: 72 vs 69 years, 45 vs 43 pack-years and 70% vs 70% male for cases and controls, respectively) and found the performance of the susceptibility score across quintiles was unchanged (OR range 1–58, p<0.01). Further studies will need to be carried out to address these issues.

It is likely that genetic susceptibility to lung cancer results from a variable combination of several genetic variants in genes encoding proteins involved in several pathways activated by chronic smoke exposure and the inflammatory response that follows. A candidate gene (ie, hypothesis-driven) approach was used to identify potentially functional SNPs associated with the development of lung cancer. Although the SNPs identified in this study may only reflect linkage disequilibrium with functional variants nearby, these SNPs are likely to have functional effects and involvement directly with susceptibility to lung cancer. Two SNPs are from genes involved in the

metabolism of smoking-derived carcinogens (N-acetyltransferase 2 and cytochrome P450 2E1) and previously linked to smoking-related cancers of the aerodigestive system.[32 33] Five SNPs are from genes encoding inflammatory cytokines implicated in carcinogenesis or lung matrix remodelling (COPD), the latter strongly implicated in lung cancer development (interleukins 1, 8 and 18, tissue necrosis factor receptor, Toll-like receptor 9).[34–42] Two SNPs are from genes that have been implicated in smoking addiction and lung cancer (dopamine D2 receptor and dopamine transporter 1).[43 44] Two SNPs are functional and found in genes involved in the antioxidant response to aero-pollutants such as smoking (α1-antichymotrypsin and extracellular superoxide dismutase).[30 31 45 46] Both of these have been associated with COPD, and one is upregulated in lung cancer. Six of the SNPs are found in genes involved in processes such as cell-cycle control, DNA repair and apoptosis, and associated with lung cancer in previously published studies (xeroderma pigmentosum complementary group D, p73, Bcl-2, FasL, Cerb1 and REV1).[47–55] Two of the SNPs are from genes encoding integrins also implicated in apoptosis, cancer susceptibility and, for one, upregulation in lung cancer cells.[56–58] One of the SNPs (α5 nAChR) has recently been associated with both lung cancer and COPD in genome-wide association studies.[16 59–62] This receptor appears to de directly related to nicotine effects on airway inflammation.[63] As can be seen, the SNP panel (table 3) is made up of a variety of SNPs from genes implicated in metabolism of smoke-derived carcinogens, oxidant response, cell-cycle control and inflammation. Twelve of these SNPs have been associated with lung cancer in other cohorts. It is likely that other SNPs from as yet unidentified genes will be identified in the future. To assess further the utility of the lung cancer susceptibility score, a prospective study is in progress. To date, the lung cancer cases (n = 43) have the same mean and distribution as the lung cancer cases reported in this study (unpublished data). Further case–control and functional studies will be needed to further explore the role of these SNPs in lung cancer susceptibility.

We propose that clinical utility of genotype data requires that many SNPs are analysed and their effects combined with other epidemiological factors of relevance.[20] The algorithm approach used in this study is comparable to that recently published for prostate cancer[20] and involves minimal assumptions (not hierarchical or path analysis based). The patient's score can be compared with the scores in smokers with least susceptibility to lung cancer (lowest quintiles) in a simple linear fashion. Such an approach is comparable to the risk tools developed by others[18 19 22 23] and similar in approach to recently published studies on risk in diabetes, where SNP data were combined with non-genetic risk variables to refine existing risk models.[64 65] The clinical utility of the lung cancer susceptibility score was assessed by ROC analysis. This showed the c statistic to be 0.79 and, at a cut-off of ⩾3, an estimated sensitivity of 89% and corresponding specificity of 45%. After FDR analysis, 12 significantly associated SNPs were included in the model, with little decrease in the AUC (0.75 vs 0.77). These findings are comparable to the ROC performance of the Framingham Score (c statistic = 0.74),[22] although other methods of assessing model performance have been advocated (eg, reclassification table approach[66]). The c statistic for the 20-SNP panel on its own was 0.68 (and 0.70 when combined with family history), indicating its utility in the current cohort. In contrast with the models for diabetes and prostate cancer, in our risk model for lung cancer it has been possible to account for the important environmental risk factor of smoking. There is evidence, although limited, that

Main messages

► Lung cancer results from the combined effects of smoking and genetic susceptibility.

► Chronic obstructive pulmonary disease is a common pre-existing and independent risk factor for lung cancer.

► Genetic susceptibility for lung cancer includes genetic variants (single nucleotide polymorphims (SNPs)) conferring reduced risk ("protective") best identified using a healthy smoking cohort.

► Genetic susceptibility for lung cancer results from the combined effects of genetic variants (SNPs) conferring either susceptibility or protective predisposition.

► Genetic and non-genetic variables can be combined to give a global risk score for susceptibility to lung cancer.

Current research questions

► Can the lung cancer susceptibility score be validated in smokers and ex-smokers in prospective studies and other populations?

► Will use of the lung cancer susceptibility score improve patient outcomes to reduce risk of lung cancer and/or detect lung cancer at a treatable stage?

genetic testing may positively alter the behaviour of smokers in the context of smoking cessation (increase intent and possibly improve quit rate[67 68]) or by lowering smoking prevalence.[69] The lung cancer susceptibility score may also have utility in early diagnosis of lung cancer where delays in diagnosis may affect survival.[70] Although further validation studies are required, this study suggests that genetic data may be combined with other risk variables from smokers or ex-smokers to identify individuals most susceptible to developing lung cancer.

REFERENCES

1. **Mattson ME,** Pollack ES, Cullen JW. What are the odds that smoking will kill you? *Am J Pub Health* 1987;**77**:425–431.
2. **Alberg AJ,** Samet JM. Epidemiology of lung cancer. *Chest* 2003;**123**:21–49.
3. **Jonsson S,** Thorsteinsdottir U, Gudbjartsson DF, et al. Familial risk of lung carcinoma in the Icelandic population. *JAMA* 2004;**292**:2977–2983.
4. **Lichtenstein P,** Holm NV, Verkasalo PK, et al. Environmental and heritable factors in the causation of cancer: analyses of cohorts of twins from Sweden, Denmark and Finland. *N Eng J Med* 2000;**343**:78–85.
5. **Young RP,** Hopkins R, Eaton TE. Forced expiratory volume in one second: not just a test of lung function but a biomarker of premature death from all causes. *Eur Respir J* 2007;**30**(4):616–622.
6. **Brody JS,** Spira A. Chronic obstructive pulmonary disease, inflammation, and lung cancer. *Proc Am Thorac Soc* 2006;**3**:535–538.
7. **Mannino DM,** Watt G, Hole D, et al. The natural history of chronic obstructive pulmonary disease. *Eur Respir J* 2006;**27**:627–643.
8. **Lokke A,** Lang P, Scharling H, et al. Developing COPD: a 25 year follow up study of the general population. *Thorax* 2006;**61**:935–939.
9. **Mannino DM,** Aguayo SM, Petty TL, et al. Low lung function and incident lung cancer in the United States: data from the first NHANES follow-up. *Arch Int Med* 2003;**163**:1475–1480.
10. **Young RP,** Hopkins RJ, Christmas T, et al. COPD prevalence is increased in lung cancer independent of age, gender and smoking history. *Eur Respir J* 2009; 5th Feb on-line.
11. **Anthonisen NR.** Prognosis in chronic obstructive pulmonary disease: results from multicenter clinical trials. *Am Rev Respir Dis* 1989;**140**:S95–S99.
12. **Xu H,** Spitz MR, Amos CI, et al. Complex segregation analysis reveals a multigene model for lung cancer. *Hum Genet* 2005;**116**:121–127.
13. **Shields P,** Harris C. Cancer risk and low-penetrance susceptibility genes in gene-environment interactions. *J Clin Oncol* 2000;**18**:2309–2315.
14. **Zhou W,** Liu G, Park S, et al. Gene-smoking interaction associations for the ERRC1 polymorphisms in the risk of lung cancer. *Cancer Epidemiol Biomarkers Prev* 2005;**14**:491–496.
15. **Schwartz AG,** Prysak GM, Bock CH, et al. The molecular epidemiology of lung cancer. *Carcinogenesis* 2007;**28**:507–518.
16. **Young RP,** Hopkins RJ, Hay BA, et al. Lung cancer gene associated with COPD: triple whammy or confounding effect? *Eur Respir J* 2008;**32**:1158–1164.
17. **Moskvina V,** Holmans P, Schmidt KM, et al. Design of case-controls studies with unscreened controls. *Ann Hum Genet* 2005;**69**:566–576.
18. **Cassidy A,** Duffy SW, Myles JP, et al. Lung cancer risk prediction: a tool for early detection. *Int J Cancer* 2006;**120**:1–6.
19. **Spitz MR,** Hong WK, Amos CI, et al. A risk model for prediction of lung cancer. *J Natl Cancer Inst* 2007;**99**(9):715–726
20. **Zheng SL,** Sun J, Wiklund F, et al. Cumulative association of five genetic variants with prostate cancer. *NEJM* 2008;**358**:910–919.
21. **Lee K,** Koval JJ. Determinants of the best significance level in forward stepwise regression. *Commun Stats* 1997;**26**:559–575.
22. **Grundy SM,** Balady GJ, Criqui MH, et al. Primary prevention of coronary heart disease: guidance from Framingham. *Circulation* 1998;**97**:1876–1887.
23. **Gail MH,** Brinton LA, Byar DP, et al. Projecting individual probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst* 1989;**81**:879–1886.
24. **Neugut AI,** Jacobson JS. Women and lung cancer: Gender equality at a crossroad? *JAMA* 2006;**296**:218–219.
25. **Easton DF,** Peto J, Babiker AG, et al. Floating absolute risk: an alternative to relative risk in survival and case-control analysis avoiding an arbitrary reference group. *Statistics in Medicine* 1991;**10**:1025–1035.
26. **Plummer M.** Improved estimates of floating absolute risk. *Statistics in Medicine* 2004;**23**:93–104.
27. **Pritchard JK,** Stephens M, Donnelly P. Inference of population structure from multi-locus genotype data. *Genetics* 2000;**155**:945–959
28. **Yang P,** Allen MS, Aubry MC, et al. Clinical features of 5,628 primary lung cancer patients; experience at Mayo clinic from 1997 to 2003. *Chest* 2005;**128**:452–462.
29. **Wald NJ,** Hackshaw AK, Frost CD. When can a risk factor be used as a worthwhile screening test? *Brit Med J* 1999;**319**:1562–1565.
30. **Young RP,** Hopkins RJ, Black PN, et al. Functional variants of anti-oxidant genes in smokers with COPD and in those with normal lung function. *Thorax* 2006;**30**:1–7.
31. **Juul K,** Tybjærg-Hansen A, Marklund S, et al. Genetically increased antioxidative protection and decreased COPD. *Am J Respir Crit Care Med* 2006;**173**:858–864.
32. **Wikman H,** Thiel S, Jäger B, et al. Relevance of N-acetyltransferase 1 and 2 (NAT1,NAT2) genetic polymorphisms in non-small cell lung cancer susceptibility. *Pharmacogenetics* 2001;**11**:157–168.
33. **Wu X,** Amos CI, Kemp BL, et al. Cytochrome P450 2E1 Dra I polymorphisms in lung cancer in minority populations. *Cancer Epidemiol Biomarkers Prev* 1998;**7**:13–18.
34. **Engels EA,** Wu X, Gu J, et al. Systematic evaluation of genetic variants in the inflammation pathway and risk of lung cancer. *Cancer Res* 2007;**67**:6520–6527.
35. **Zienolddiny S,** Ryberg D, Maggini V, et al. Polymorphisms of the interleukin-1 B gene are associated with increased risk of non-small cell lung cancer. *Int J Cancer* 2004;**109**:353–356.
36. **Campa D,** Zienolddiny S, Maggini V, et al. Association of a common polymorphism in the cyclooxygenase 2 gene with risk of non-small cell lung cancer. *Carcinogenesis* 2004;**25**:229–235.
37. **Hoshino T,** Kato S, Oka N, et al. Pulmonary inflammation and emphysema: role of the cytokines IL-18 and IL-13. *Am J Respir Crit Care Med* 2007;**176**:49–62.
38. **Hodge SJ,** Hodge GL, Reynolds PN, et al. Increased production of TGFB and apoptosis of t lymphocytes isolated from peripheral blood in COPD. *Am J Physiol Lung Cell Mol Physiol* 2003;**285**:L492–L499.
39. **Bhavsar TM,** Cerreta JM, Cantor JO, et al. Short term cigarette smoke exposure predisposes the lung to secondary injury. *Lung* 2007;**185**:227–233.
40. **Droemann D,** Albrecht D, Gerdes J, et al. Human lung cancer cells express functionally active Toll-like Receptor 9. *Respiratory Research* 2005;**6**:1–10.
41. **Noakes PS,** Hale J, Thomas R, et al. Maternal smoking is associated with impaired neonatal toll-like receptor mediated immune responses. *Eur Respir J* 2006;**28**:721–729.
42. **Lazarus R,** Klimecki WT, Raby BA, et al. Single nucleotide polymorphisms in the toll-like receptor 9 gene: frequencies, pairwise linkage disequilibrium and haplotypes in three US ethnic groups and exploratory case-control disease association studies. *Genomics* 2003;**81**:85–91.
43. **Campa D,** Zienolddiny S, Lind H, et al. Polymorphisms of dopamine receptor/transporter genes and risk of non-small cell lung cancer. *Lung Cancer* 2006;**56**(1):17–23.
44. **Wu X,** Hudmon KS, Detry MA, et al. D2 dopamine receptor gene polymorphisms among African-Americans and Mexican-Americans: a lung cancer case-control study. *Cancer Epidemiol Bio Prevent* 2000;**9**:1021–1026.

45. **Ishii T,** Matsuse T, Teramoto S, et al. Association between alpha1-antichymotrypsin polymorphisms and susceptibility to chronic obstructive pulmonary disease. *Eur J Clin Invest* 2000;**30**:543–548.

46. **Zelvyte I,** Wallmark A, Piitulainen E, et al. Increased plasma levels of serine proteinase inhibitors in lung cancer patients. *Anticancer Research* 2004;**24**:241–247.

47. **Hu Z,** Wei Q, Wang X, et al. DNA repair gene XPD polymorphism and lung cancer risk: a meta-analysis. *Lung Cancer* 2004;**46**:1–10.

48. **Yin JY,** Vogel U, Ma Y, et al. A haplotype encompassing the variant allele of DNA repair gene polymorphism ERCC2/XPD Lys 751Gln but not the variant allele of Asp312Asn is associated with risk of lung cancer in a northeastern Chinese population. *Canc Genet Cytogen* 2007;**175**:47–51.

49. **Spitz MR,** Wu X, Wang Y, et al. Modulation of nucleotide Excision repair capacity by XPD polymorphisms in lung cancer patients. *Cancer Res* 2001;**61**:1354–1357.

50. **Goujun L,** Wang LE, Chamberlain RM, et al. p73 G4C14-to-A4T14 polymorphism and risk of lung cancer. *Can Res* 2004;**64**:6863–6866.

51. **Sata M,** Takabatake N, Inoue S, et al. Intronic single nucleotide polymorphisms in Bcl-2 are associated with chronic obstructive pulmonary disease severity. *Respirology* 2007;**12**:43–41.

52. **Jain M,** Kumar S, Lal P, et al. Role of Bcl-2 (ala43thr), CCND1 (G870A) and FAS (A-670G) polymorphisms in modulating the risk of developing esophageal cancer. *Cancer Detec Prev* 2007;**31**(3):225–232.

53. **Zhang X,** Miao X, Sun T, et al. Functional polymorphisms in cell death pathway genes FAS and FASL contribute to risk of lung cancer. *J Med Genet* 2005;**42**:479–484.

54. **Sakiyama T,** Kohno T, Mimaki S, et al. Association of amino acid substitution polymorphisms in DNA repair genes TP53, POLI, REV1, and LIG4 with lung cancer. *Int J Cancer* 2005;**114**:730–737.

55. **Rudd MF,** Webb EL, Matakidou A, et al. Variants in the GH-IGF axis confer susceptibility to the lung cancer. *Genome Res* 2007; August 693–701.

56. **Jakubowska A,** Gronwald J, Menkiszak J, et al. Integrin beta3 Leu 33Pro polymorphism increases BRCA1-associated ovarian cancer risk. *J Med Genet* 2007;**44**:408–411.

57. **Zhu CQ,** Popova SN, Brown ER, et al. Integrin alpha11 regulates IGF2 expression in fibroblasts to enhance tumorigenicity of human non-small-cell lung cancer cells. *Proc Natl Acad Sci* 2007;**104**:11754–11759.

58. **Chong IW,** Chang MY, Chang HC, et al. Great potential of a panel of multiple hMTH1, SPD, ITGA11, and COL11A1 markers for the diagnosis of patients with non-small cell lung cancer. *Oncol Rep* 2006;**16**:981–988.

59. **Thorgeirsson TE,** Geller F, Sulem P, et al. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature* 2008;**452**:638–642.

60. **Hung RJ,** mckay JD, Gaborieau V, et al. A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature* 2008;**452**:633–637.

61. **Amos CI,** Wu X, Broderick P, et al. Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nat Genetics* 2008; April 2nd, 1–7.

62. **Pillai SG,** Shianna K, Ge D, et al. Genome-wide association study of chronic obstructive pulmonary disease (COPD) in a case control population from Norway [abstract]. American Thoracic Society International Conference: 16-26 May 2008: Toronto pg A776.

63. **Gwilt CR,** Donnelly LE, Rogers DF. The non-neuronal cholinergic system in the airways: an unappreciated regulatory role in pulmonary inflammation? *Pharmacol Therapeut* 2007;**115**:208–222.

64. **Lyssenko V,** Jonsson A, Almgren P, et al. Clinical risk factors, DNA variants, and the development of type 2 diabetes. *NEJM* 2008;**359**:2220–2232.

65. **Meigs JB,** Shrader P, Sullivan LM, et al. Genotype score in addition to common risk factors for prediction of type 2 diabetes. *NEJM* 2008;**359**:2208–2219.

66. **Cook NR.** Use and misuse of the receiver operator characteristic curve in risk prediction. *Circulation* 2007;**115**:928–935.

67. **McBride CM,** Bepler G, Lipkus IM, et al. Incorporating genetic susceptibility feedback into a smoking cessation program for African-American smokers with low income. *Cancer Epidemiol Biomarkers Prev* 2002;**11**:521–528.

68. **Lerman C,** Gold K, Audrain J, et al. Incorporating biomarkers of exposure and genetic susceptibility into smoking cessation treatment: effects on smoking-related cognitions, emotions and behavior change. *Health Psychology* 1997;**16**:87–99.

69. **Strange C,** Dickson R, Carter C, et al. Genetic testing for alpha1-antitrypsin deficiency. *Genet Med* 2004;**6**:240–210.

70. **Hamilton W,** Peters TJ, Round A, et al. What are the clinical features of lung cancer before the diagnosis is made? A population based case-control study. *Thorax* 2005;**60**:1059–1065.