



Chromosome 4q31 locus in COPD is also associated with lung cancer

R.P. Young^{*,#}, C.F. Whittington^{*}, R.J. Hopkins^{*}, B.A. Hay^{*}, M.J. Epton[†],
P.N. Black^{*,†} and G.D. Gamble^{*}

ABSTRACT: Chronic obstructive pulmonary disease (COPD) is the single greatest risk factor for lung cancer in smokers and is found in 50–90% of lung cancer cases. The link between COPD and lung cancer may stem in part from the matrix remodelling and repair processes underlying COPD, and the development of epithelial–mesenchymal transition (EMT) that underlies lung carcinogenesis. The Hedgehog-interacting protein (HHIP), which mediates the epithelial response (EMT) to smoking, has been implicated in COPD and lung cancer. Recent genome-wide and candidate gene studies of COPD implicate genetic variants on the chromosomal 4q31 (*HHIP*/glycophorin A (*GYP*A)) locus.

In a case–control study of smokers with normal lung function, COPD and lung cancer (subphenotyped for COPD), we show the GG genotype of the rs 1489759 *HHIP* single-nucleotide polymorphism (SNP) and the CC genotype of the rs 2202507 *GYP*A SNP confers a “protective” effect on COPD (OR 0.59, $p=0.006$ for *HHIP* and OR=0.65, $p=0.006$ for *GYP*A) and lung cancer (OR=0.70 ($p=0.05$) for *HHIP* and OR 0.70 ($p=0.02$) for *GYP*A).

This study suggests that, in smokers, genetic variants of the 4q31 locus conferring a protective effect for COPD are also protective in lung cancer. We conclude that genetic susceptibility to lung cancer includes COPD-related gene variants.

KEYWORDS: Association study, chronic obstructive pulmonary disease, Hedgehog-interacting protein, lung cancer, polymorphism

Chronic obstructive pulmonary disease (COPD) and lung cancer are lung diseases that result from the combined effects of smoking exposure and genetic susceptibility [1, 2]. Epidemiological studies show that ~20–30% of smokers develop COPD, while 10–15% develop lung cancer [3, 4]. We and others have shown that COPD confers a four- to six-fold increased risk of lung cancer compared to smokers with normal lung function [5–7], and that 50–90% of those with lung cancer have pre-existing COPD [5, 8–10]. The latter compares with 15% in randomly selected, community-based smoking controls [5]. The heritability of COPD and lung cancer is estimated to be 40–77% and 15–25%, respectively [11, 12]. These findings suggest some smokers susceptible to COPD might also be susceptible to lung cancer and that some of the genetic factors conferring this dual susceptibility might overlap [13–15]. Recent interest in lung cancer has focused on the role of smoking in epithelial–mesenchymal transition (EMT) [16], a process driven by release of growth factors and matrix metalloproteinases during lung remodelling and repair [17, 18]. The latter

underlies the development of COPD and might link lung cancer with COPD at a pathological level, as EMT is known to initiate and promote lung carcinogenesis [17–19].

Over the past 2 yrs, genome-wide association (GWA) studies in lung cancer and COPD have reported significant associations at several chromosomal loci, some of which overlap [20–27]. The presence of overlapping loci suggest common pathogenetic pathways may underlie these pulmonary complications of smoking (table 1). This is analogous to the interrelated pathways underlying obesity and type 2 diabetes, in which the *FTO* (fat mass and obesity-associated) gene has been implicated in both diseases [28]. In the context of identifying lung cancer susceptibility genes, the above observations have several important epidemiological and clinical implications. First, that lung cancer is a complex phenotype that includes COPD, a disease characterised by a reduced forced expiratory volume in 1 s (FEV₁) secondary to a variable combination of parenchymal destruction (emphysema) and small airways disease. In the absence of spirometry, 50–80% of

AFFILIATIONS

^{*}Dept of Medicine, Auckland Hospital,
[#]Synergz Biosciences Ltd, Auckland, and
[†]Dept of Medicine, University of Otago, Christchurch, New Zealand.

CORRESPONDENCE

R. Young
Dept of Medicine
Auckland Hospital
Private Bag 92019
Auckland
New Zealand
E-mail: roberty@adhb.govt.nz

Received:
March 02 2010
Accepted after revision:
May 14 2010

COPD cases remain undiagnosed [29]. In order to better explore associations between single nucleotide polymorphisms (SNPs), COPD and/or lung cancer, the COPD phenotype requires identification through spirometry assessment [13, 14]. Secondly, to better examine SNP associations with lung cancer and/or COPD, three study populations should be employed: smokers with normal lung function (resistant smokers with no evidence of COPD); smokers with COPD; and smokers with lung cancer in whom spirometry has been performed to identify coexisting COPD [13, 14]. Using such an approach, we identified that the chromosome 15q25 locus, originally associated with lung cancer in GWA studies [20–22], is also associated with COPD [30]. This observation has been replicated in subsequent GWA and candidate gene studies [23, 31]. Thirdly, the cohorts should be carefully matched for smoking history and other relevant variables, such as age, ethnicity and sex [32, 33]. Matching for smoking is particularly important, as the penetrance of SNP effects in COPD and lung cancer, reflected in the odds ratio, are likely to be related to the degree and/or duration of smoking exposure [24, 34], as it is in α_1 -antitrypsin deficiency [35]. Case-control studies, in which never-smokers have been preferentially included in the controls, may miss a clinically relevant gene by environmental (smoking) association through a dilution effect. This is analogous to a pharmacogenetic study where responsiveness to a drug is compared to genetic variants but some subjects have not taken the drug [20–22]. We propose the epidemiological data, heritability estimates and GWA study results suggest that some of the genetic susceptibility to lung cancer comes from genes associated with susceptibility to COPD.

Three GWA studies [23–25], a GWA meta-analysis [26] and a recent candidate gene study [34] have demonstrated that the

4q31 locus, which contains the Hedgehog (Hh)-interacting protein (*HHIP*) gene, is associated with COPD and lung cancer (table 1). The association of the 4q31 locus with lung cancer was reported by only one of three GWA studies [20], although the findings were inconsistent between cohorts. *HHIP* is a regulator of the Hh signalling pathway, which has been shown to be vital for embryonic lung development and is also involved in mature airway epithelial repair [36–40]. Alterations of the *HHIP* protein or its expression may lead to changes in lung repair mechanisms, supporting a role in the development of COPD. In addition to this, the Hh signalling pathway is involved in EMT, mediates cigarette smoke-induced oncogenic transformation of bronchial epithelial cells and is necessary for cellular proliferation of many lung cancer cell lines [16, 40, 41]. As there is significant support for the role of *HHIP* and the Hh signalling pathway in the development of both COPD and lung cancer, we investigated the role of the 4q31 locus in our case-control study of smokers with and without COPD, and smokers with lung cancer in whom spirometry was performed to allow subphenotyping for coexisting COPD.

MATERIALS AND METHODS

Study subjects

Subjects recruited were of Caucasian ancestry based on their grandparents' descent (all four grandparents of Caucasian descent). Patients with lung cancer were identified through hospital records and specialist clinics between 2004 and 2007. Lung cancer cases were aged >40 yrs and their diagnosis confirmed through histological or cytological specimens in 95% of cases. Nonsmokers with lung cancer were excluded from the study and only primary lung cancer cases with the following pathological diagnosis were included: adenocarcinoma, squamous cell cancer, small cell cancer and nonsmall cell cancer (generally large cell or bronchoalveolar subtypes). Lung function measurement (pre-bronchodilator) was performed within 3 months of lung cancer diagnosis, prior to surgery and in the absence of pleural effusions or lung collapse on plain chest radiographs. Lung function conformed to American Thoracic Society (ATS) standards for reproducibility, with the highest value of the best three acceptable blows used for classification of COPD status. For lung cancer cases that had already undergone surgery, pre-operative lung function tests performed by the lung function laboratory (Auckland Hospital, Auckland, New Zealand) was sourced from medical records.

Patients with COPD were identified through hospital specialist clinics between 2003 and 2007. Subjects recruited into the study were aged 40–80 yrs, with a minimum smoking history of 20 pack-yrs and COPD confirmed by a respiratory specialist based on pre-bronchodilator spirometric criteria. Control subjects were recruited based on the following criteria: age 45–80 yrs and with a minimum smoking history of 20 pack-yrs. Control subjects were volunteers who were identified through either a community postal advert or while attending community-based retired military servicemen's clubs located in the same patient catchment area as those attending the lung cancer and COPD hospital clinics. All participants gave written, informed consent and underwent blood sampling for DNA extraction, spirometry and an investigator-administered

TABLE 1 Chromosomal loci associated with chronic obstructive pulmonary disease (COPD) and forced expiratory volume in 1 s (FEV₁) and lung cancer from genome-wide association studies

Disease	Chromosomal loci	Candidate genes	[Ref.] [#]
COPD [‡] and FEV ₁	15q25	<i>CHRNA3/5</i>	[23]
	4q31	<i>HHIP, GYP A</i>	[23–26]
	6p21	<i>BAT3/AGER</i>	[25, 26]
	5q33	<i>HTR4</i>	[25, 26]
	15q11	<i>SNRPN</i>	[27]
	1q21	<i>CRP/IL6R</i>	[27]
	4q24	<i>GSTCD</i>	[25, 26]
Lung cancer	15q25	<i>CHRNA3/5</i>	[20–22]
	4q31	<i>GYP A*</i>	[20]
	5p15	<i>TERT/CRR9</i>	[2, 20, 21]
	1q21-23	<i>CRP/IL6R</i>	[20]
	6p21	<i>BAT3</i>	[2, 20, 21]
	3q28	<i>IL1RAP</i>	[20]

[#]: available at www.genome.gov/gwastudies (last accessed February 25, 2010); [‡]: defined on the basis of forced expiratory volume in 1 s measurements; ^{*}: did not reach genome-wide significance levels.

questionnaire. Controls with COPD based on spirometry (Global Initiative for Chronic Obstructive Lung Disease (GOLD) stage I or higher) were excluded from further analysis.

Spirometry was performed using a portable spirometer (EasyOne™; ndd Medizintechnik AG, Zurich, Switzerland). COPD was defined according to GOLD criteria (stage II or higher; FEV₁/forced vital capacity (FVC) <70% and FEV₁ ≤80% predicted) using pre-bronchodilator spirometric measurements. A modified ATS respiratory questionnaire was administered to all cases and controls, which collected data on demographic variables, such as age, sex, medical history, family history of lung disease, active and passive tobacco exposure, respiratory symptoms and occupational aeropollutant exposures. The study was approved by the Multi Centre Ethics Committee (Wellington, New Zealand).

Study design

The present cross-sectional case-control study compared smokers of the same ethnicity and comparable demographic variables (specifically age, sex and smoking history). The controls in the current study were carefully chosen to be representative of the majority of smokers who have maintained normal or near-normal lung function despite decades of smoking [3, 42, 43]. Accordingly, such a group best reflects those smokers least likely to acquire lung cancer or COPD, minimising phenotype misclassification and improving the power to detect differences between affected and unaffected smokers [44].

Genotyping

Genomic DNA was extracted from whole-blood samples using standard salt-based methods and purified genomic DNA was aliquoted (10 ng·μL⁻¹) into 96-well plates. Samples were genotyped using Taqman® SNP genotyping assays (Applied Biosystems, Carlsbad, CA, USA) utilising minor groove-binding probes. Assays were run in 384-well plates according to the manufacturer's instructions. PCR cycling was performed on both GeneAmp® PCR System 9700 and 7900HT Fast Real-Time PCR System (Applied Biosystems) devices. Real-time amplification plots of selected plates were used to verify endpoint allelic discrimination to establish reliability. The present study investigated the genotype frequencies of two SNPs in the 4q31 region: rs2202507 (glycophorin A (*GYP*A), assay ID C_1509159_10; Applied Biosystems), which was identified in a GWA study on lung cancer [20], and rs1489759 (*HHIP*, assay ID C_8977401_10; Applied Biosystems), which is situated 93 kb upstream of *HHIP* and representative of the SNP cluster identified by previous COPD GWA related studies [23–26] (linkage disequilibrium (LD) for $D' = 1.0$, $r^2 \geq 0.93$ with rs12504628, rs13147758, rs1828591, rs1980057 and rs13118928; fig. 1b). The rs2202507 SNP lies ~200 kb downstream from the *GYP*A gene and 300 kb upstream of the *HHIP* gene, yet exhibits some LD with the cluster of SNPs including rs1489759 ($D' = 0.70$, $r^2 = 0.41$). The two SNPs are found at either end of a 290-kb LD block (fig. 1a). Call rates of $\geq 97\%$ for each SNP in each cohort were achieved.

Analysis

Patient characteristics in the cases and controls were compared by ANOVA for continuous variables and Chi-squared test for

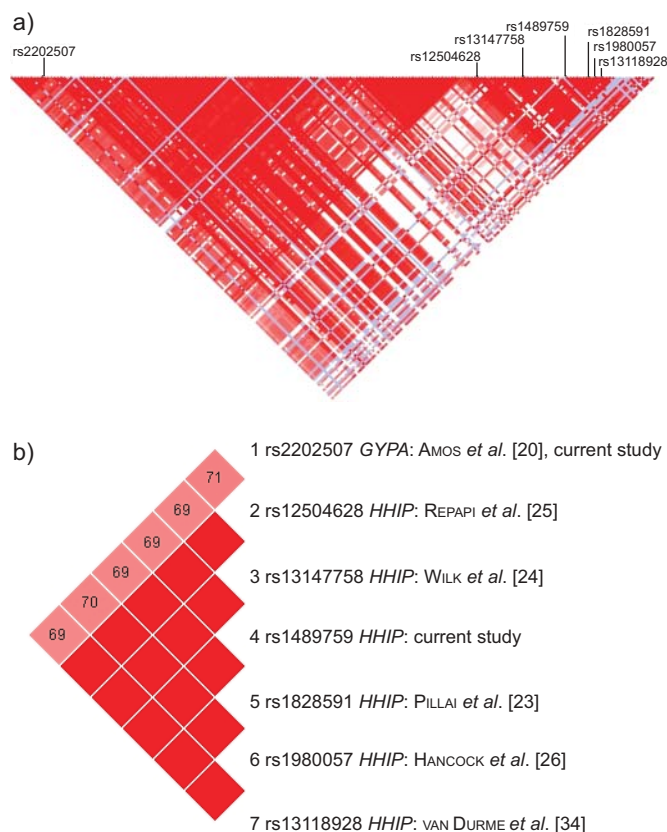


FIGURE 1. Linkage disequilibrium map of the chromosome 4q31 region including the Hedgehog-interacting protein (*HHIP*) and glycophorin A (*GYP*A) single-nucleotide polymorphisms.

discrete variables (Mantel–Haenszel, odds ratio). Genotype and allele frequencies were checked for each SNP by Hardy–Weinberg Equilibrium (HWE). Population admixture across cohorts was performed using structure analysis on genotyping data from 40 unrelated SNPs [45]. Distortions in the genotype and allele frequencies were identified between cases and controls using two-by-two contingency tables. Both the additive (allelic) and genotype-based genetic models were tested, although the latter is preferred [46].

RESULTS

Demographic variables and genotyping

Characteristics of the lung cancer cases, COPD cases and healthy control smokers are summarised in table 2. The demographic variables and histological subtypes of the lung cancer cases are comparable to previously published data [47]. The staging at diagnosis was also comparable to this published series (data not shown), suggesting the lung cancer cohort is representative. The COPD cases have higher pack-yr exposure than the lung cancer cases and healthy control smokers ($p < 0.05$). This reflects outliers with high smoking histories in the COPD cohort and no difference exists after log transformation of pack-yrs (data not shown). All cohorts were comparable with respect to age started smoking, years smoked, years since cessation of smoking and cigarettes smoked per day (table 2). Overall, we believe the three groups had comparable smoking exposure. The lower frequency of current smokers in the lung

TABLE 2 Summary of characteristics for the lung cancer, chronic obstructive pulmonary disease (COPD) and control smokers

Parameter	Lung cancer	COPD	Control smokers
Subjects n	454	458	488
Males %	53	59	60
Age yrs	69±10	66±9	65±10
Height* m	1.67±0.08	1.68±0.09	1.69±0.09
Smoking history			
Current smoking %	35	40	48
Age started yrs	18±4	17±3	17±3
Smoking exposure yrs	41±12	42±11	35±11
Smoking exposure* pack-yrs	41±25	47±20*	40±19
Smoking exposure cigarettes·day ⁻¹	20±10	23±9	24±11
Time since smoking cessation yrs	11.4±6.7	9.8±7.4	13.9±8.1
History of other exposures %			
Work dust exposure*	63	59	47
Work fume exposure	41	40	38
Asbestos exposure*	23	22	16
Family history %			
COPD	33	37	28
Lung cancer*	19	11	9
Lung function			
FEV ₁ * L	1.86±0.48	1.25±0.48	2.86±0.68
FEV ₁ * % pred	73	46	99
FEV ₁ /FVC* %	64±13	46±8	78±7
Spirometric COPD ^{#,*} %	51	100	0

Data are presented as mean±SD, unless otherwise stated. FEV₁: forced expiratory volume in 1 s; % pred: % predicted; FVC: forced vital capacity. #: according to Global Initiative for Chronic Obstructive Lung Disease criteria, stage II or higher; *: no significant difference after log transformation due to skewed distribution. *: p<0.05.

cancer and COPD cohorts compared to the healthy smoker group (35, 40 and 48%, respectively), probably reflects the presence of symptoms from pulmonary complications of smoking (primarily breathlessness from COPD) stimulating higher smoking cessation rates. The lung cancer cases, COPD cases and smoking controls were also comparable with respect to other aeropollutant exposures, although the lung cancer and COPD cohorts had higher asbestos exposure (23, 22 and 16%, respectively), slightly higher dust exposure (63, 59 and 47%, respectively) but comparable fume exposure (41, 40 and 38%, respectively). The lung cancer cohort reported higher rates of a family history of lung cancer compared to the COPD cases and healthy smokers (19, 11 and 9%, respectively), whereas family history of COPD was marginally higher in those with COPD (33, 37 and 28%, respectively). Mean height was slightly lower in lung cancer cases compared to control subjects (p<0.05) but no different after adjustment for sex differences. As expected, lung function was worse in the lung cancer and COPD cohorts compared with the healthy smoker controls. Testing of lung function (as described above) was achieved in 94% of lung cancer cases and allows stratification of results to test for an interactive or confounding effect of COPD.

Genotype frequencies for the rs1489759 and rs2202507 SNPs are shown in tables 3 and 4, respectively. The genotype frequencies were comparable to those reported in the literature and from the International Hapmap Project (www.hapmap.org). The observed genotypes for the two SNPs in this study were 65% concordant, illustrating the LD over this large physical distance. As both SNPs were in HWE and amplification plots were used to ensure correct genotype calls, significant genotyping error can be excluded. We found no evidence for population stratification between the cohorts using 40 unlinked SNPs from unrelated genes (mean Chi-squared 3.3; p=0.58) [30, 45].

Comparison with smoking controls

Those subjects who were homozygotic for the minor allele (GG genotype) of the rs1489759 *HHIP* SNP were found to be more prevalent in the control group (17%) versus both the COPD (11%, OR 0.59; p=0.006) and lung cancer (13%, OR 0.70; p=0.051) groups (table 3). Similarly, those homozygotic for the minor allele (CC genotype) of the rs2202507 *GYP A* SNP were more prevalent in the resistant smoker group (27%) compared to those in the COPD (19%, OR 0.65; p=0.006) and lung cancer (21%, OR 0.70; p=0.023) groups (table 4). Distortion in allele frequency showed a similar result, with reduced *HHIP* G allele and reduced *GYP A* C allele in the diseased cohorts compared with controls, but tended to be less significant. There was no support for an additive model on formal testing. When the lung cancer cases were stratified by available spirometry data (>94% of lung cancer cases; n=419 and n=416 for *HHIP* and *GYP A* genotyping, respectively), into those with and without COPD (according to GOLD criteria, stage II or higher), the distribution of minor allele homozygotes for both SNPs does not change significantly (for *HHIP*, 11 and 13%, respectively, compared with 17% in controls; for *GYP A*, 20 and 19%, respectively, compared with 27% in controls). The effect sizes of the homozygosity for the minor allele in these subanalyses remain the same (OR range 0.61–0.73), although the p-values are degraded due to smaller sample sizes. When stratification of the lung cancer cohort is performed according to GOLD stage I or higher (to specify COPD phenotype), genotype frequencies were unchanged. Further subanalysis was performed by grouping all subjects with COPD based on spirometry (including COPD only and lung cancer with COPD groups). The resulting protective effect (tables 3 and 4) was nearly identical to that when using the COPD cohort alone, although combining the cohorts achieved greater significance (OR 0.60 (p=0.003) and OR 0.66 (p=0.004) for the *HHIP* and *GYP A*, respectively). This suggests that the SNP associations are not sensitive to spirometry-based criteria for COPD and that the 4q31 locus is associated with reduced lung cancer, even in the absence of COPD. In a multivariate analysis, no associations were found between SNP genotype and height, lung cancer histology, FEV₁ % predicted or FEV₁/FVC ratio (table 5). We found that the SNP genotype effect was independent of age, sex and smoking history.

DISCUSSION

The present case-control study confirms the association of the 4q31 locus with COPD, as described by several recent GWA studies and a candidate gene study [23–26, 34]. To our knowledge, this is the first study to confirm the association

TABLE 3 Genotype frequencies for the rs1489759 Hedgehog-interacting protein single-nucleotide polymorphism in chronic obstructive pulmonary disease (COPD) and lung cancer (LC) (total and subgrouped by pre-operative lung function) cohorts compared with smoking controls

	Subjects	A	G	OR (95% CI) [#]	p-value [#]	AA	AG	GG	OR (95% CI) [†]	p-value [†]
Primary cohorts										
Controls	484 (99)	579 (60)	389 (40)			178 (37)	223 (46)	83 (17)		
COPD	457 (99)	594 (65)	320 (35)	0.80 (0.66–0.97)	0.002	187 (41)	220 (48)	50 (11)	0.59 (0.40–0.88)	0.006
LC	445 (99)	563 (63)	327 (37)	0.86 (0.71–1.05)	0.127	174 (39)	215 (48)	56 (13)	0.70 (0.47–1.02)	0.051
LC subgroup analyses										
LC with COPD	213	292 (69)	134 (31)	0.68 (0.53–0.88)	0.002	103 (48)	86 (40)	24 (11)	0.61 (0.39–1.02)	0.047
LC only	206	276 (67)	136 (33)	0.73 (0.57–0.94)	0.01	92 (47)	82 (40)	27 (13)	0.73 (0.44–1.19)	0.185
COPD subgroup analysis										
COPD and LC with COPD	670	869 (65)	471 (35)	0.81 (0.68–0.96)	0.014	273 (41)	323 (48)	74 (11)	0.60 (0.42–0.85)	0.003

Data are presented as n or n (call rate %), unless otherwise stated. [#]: G versus A compared with matched smoking controls (Mantel–Haenszel); [†]: GG versus AG/AA compared with matched smoking controls (Mantel–Haenszel).

between the 4q31 locus and lung cancer [20], supporting the possible “protective” role of the 4q31 locus (and possibly the *HHIP* gene) in the development of lung cancer, as well as COPD. The minor allele homozygotic genotype of both the rs1489759 (*HHIP*) and rs2202507 (*GYPA*) SNPs is associated with a significantly decreased risk of both COPD and lung cancer. Stratification of the lung cancer cohort into those with and without COPD (according to GOLD spirometry criteria) produced similar results, excluding an interactive or confounding effect from coexisting COPD. Due to the smaller sample size, statistical significance is lost, although genotype frequency and effect size are unchanged. This study also shows that the *HHIP* locus has a protective role in lung cancer with and without coexisting COPD, highlighting the dual role genetic variants may have in these closely related smoking-induced lung diseases.

This study supports not only the findings of previous GWA and candidate gene studies that have confirmed the association of the 4q31 locus with COPD and lung cancer, but also demonstrates that the candidate SNPs we have chosen have an equal effect size on both diseases. The majority of the interest in this locus has come from studies of COPD and lung function. All of the SNPs previously identified upstream of the *HHIP* gene are in tight LD with our chosen SNP, rs1489759, and share similar allele frequencies ($D' = 1.0$, $r^2 \geq 0.93$; minor allele frequency 0.441–0.425 according to HapMap; fig. 1). We have shown that the minor allele homozygotic genotype is associated with “protection” (OR <1) from both lung cancer and COPD at an effect size that is in agreement with previous reported studies (OR 0.7 or 30% reduction) [34]. The effect of smoking appears to influence the strength of this association between the 4q31 locus and COPD or lung function [24, 34]. The population-based GWA study by WILK *et al.* [24] found a dose–response effect between the rs13147758 *HHIP* SNP and lung function, which was strengthened when investigated only in the subgroup of ever-smokers. Likewise, the Rotterdam Study [34] showed that the effect of the rs13118928 *HHIP* SNP was strongest in the subgroup of smokers with the greatest degree of smoking exposure, far and above that of the never-smokers. This demonstrates the need for genetic

association studies of smoking-related disease to control (or stratify) for smoking exposure in case and control cohorts. In contrast to the *HHIP* GWA population studies [24–26], we did not find an association between this locus and lung function in our lung cancer cases. This may be due to several factors, including the small cohort used for this analysis, the subject disease status (lung cancer cases, of whom 51% had coexisting COPD), the difference in smoking exposure or modifying effects of other smoking-responsive genes on lung function decline, *e.g.* *CHRNA3/5* [30]. Based on our results (distortions in genotype and allele frequencies, together with analysis for an additive model), we suggest the chromosome 4q31 effect in COPD and lung cancer is most consistent with a recessive model, in agreement with VAN DURME *et al.* [34]. Although small cohort size and case–control design remain limitations of this study, important strengths of the present study are the detailed exposure histories collected and the degree to which they are matched. Due to the low penetrance of SNPs such as those investigated in the present study, it is only through controlling for possible environmental exposures that the modest effect of these genetic variants will be recognised fully. This, together with the use of a healthy smoking control group, might explain why the chromosome 4q31 locus association with lung cancer was identified in this study but in only one of three GWA studies [20]. We suggest that the protective effect of the chromosomal 4q31 SNPs identified here might be obscured in the GWA studies, due to their inclusion of controls with undiagnosed COPD and never-smokers [20–22]. Such an effect is illustrated in tables 3 and 4, where genotype frequencies in those with COPD and lung cancer are similar, but significantly different, to healthy smoking controls, in whom COPD has been excluded by spirometry. In an attempt to sufficiently power these GWA studies, there is a tendency to combine a number of heterogeneous study populations, each possessing different inclusion criteria, each with variable (unknown) prevalence of COPD and each with different smoking exposures [33, 48]. Identifying a clean phenotype for both control and disease groups should improve the ability of a case–control study to identify true genetic associations [33, 44, 48].

TABLE 4 Genotype frequencies for the rs2202507 glycoprotein A single-nucleotide polymorphism in chronic obstructive pulmonary disease (COPD) and lung cancer (LC) (total and subgrouped by pre-operative lung function) cohorts compared with smoking controls

	Subjects	A	C	OR (95% CI) [#]	p-value [#]	AA	AC	CC	OR (95% CI) [†]	p-value [†]
Primary cohorts										
Controls	480 (99)	489 (51)	471 (49)			138 (29)	213 (44)	129 (27)		
COPD	457 (99)	505 (55)	409 (45)	0.84 (0.70–1.01)	0.061	136 (30)	233 (51)	88 (19)	0.65 (0.47–0.89)	0.006
LC	439 (97)	465 (53)	413 (47)	0.92 (0.76–1.11)	0.385	116 (26)	233 (53)	90 (21)	0.70 (0.51–0.97)	0.023
LC subgroup analyses										
LC with COPD	212	231 (54)	193 (46)	0.87 (0.69–1.10)	0.224	62 (29)	107 (50)	43 (20)	0.69 (0.46–1.04)	0.065
LC only	204	217 (53)	191 (47)	0.91 (0.72–1.16)	0.446	52 (25)	113 (55)	39 (19)	0.64 (0.42–0.98)	0.031
COPD subgroup analysis										
COPD and LC with COPD	669	736 (55)	602 (45)	0.85 (0.72–1.01)	0.053	198 (30)	340 (51)	131 (20)	0.66 (0.50–0.88)	0.004

Data are presented as n or n (call rate %), unless otherwise stated. [#]: C versus A compared with matched smoking controls (Mantel–Haenszel); [†]: CC versus AC/AA compared with matched smoking controls (Mantel–Haenszel).

The Hh signalling pathway is known to have an influence on the fetal development of the lung, especially in branching, and is involved in the complex regulation of cellular differentiation and proliferation. HHIP is a transmembrane glycoprotein that binds to and regulates the action of all three mammalian Hh ligands (Sonic hedgehog (Shh), Indian hedgehog and Desert hedgehog) [16, 36]. It is thought that HHIP is responsible for regulating the delicate balance between Shh and fibroblast growth factor (Fgf)10 signalling in the developing lung [36, 38]. Knockout of *HHIP* in mice leads to an inhibition of lung bud branching due to unopposed Shh signalling and leads to neonatal death due to respiratory failure. The disruption of branching morphogenesis is caused by the over action of Shh and the subsequent inhibition of Fgf10 [37, 38]. Additionally, Hh signalling has been shown to promote the differentiation of distal lung mesenchyme into smooth muscle cells [36]. With regards to COPD, the Hh pathway could be involved in

formation of the characteristic structural defects through inappropriate cell growth and aberrant control of smooth muscle differentiation. In addition to its role in development and cell differentiation in the embryonic lung [36], the Hh signalling pathway is known to play a role in the development of some cancers, including basal cell carcinoma and lung cancer [40, 41, 49]. WATKINS *et al.* [40] demonstrated that the Hh pathway is extensively activated during acute lung injury repair and is required for the maintenance of the malignant phenotype of small cell lung cancer (SCLC) cell lines. It has also been shown that active Hh signalling is common in SCLC tumour samples [49]. In addition, the Hh pathway has been linked to smoking-related lung cancer development [41]. By pharmacologically inhibiting the pathway, human primary bronchial epithelial cells lost the ability to form tumours in nude mice following 8 days of cigarette smoke exposure [41]. EMT is the process by which cells switch from epithelial phenotypes, which are often highly differentiated and characterised by a high level of cell–cell interactions, to mesenchymal phenotypes, which are less differentiated, have significantly fewer cell–cell interactions and can be motile [17–19]. EMT, which promotes motility, invasion and malignant transformation of cells, has been identified in patients with COPD [19] and linked to the Hh signalling pathway through its interaction with cigarette smoke [16, 41]. Although the effect of the SNPs identified upstream of the *HHIP* gene are unknown, it can be postulated that a particular form of the HHIP protein or transcriptional alteration, through its' interaction with the Hh signalling pathway, may affect the way in which the lung responds to chronic exposure to cigarette smoke.

The *HHIP* gene is not the only candidate gene explaining the 4q31 association with COPD and lung cancer. There is a putative gene, *CR620567*, located on the opposite strand to *HHIP* and running in the opposite direction, which is situated around 100 kb from the rs1489759 *HHIP* SNP. Situated near the region of interest, there are three overlapping expressed sequence tags (ESTs) that have been cloned previously from a human fetal lung library [24]. An EST is a portion of transcribed DNA often used to identify the position of an

TABLE 5 Relationship between Hedgehog-interacting protein (*HHIP*) and glycoprotein A (*GYPA*) genotype and lung function in the lung cancer cohort

Gene (SNP)	FEV ₁ L	FEV ₁ % pred	FEV ₁ /FVC
<i>HHIP</i> (rs1489759)			
AA	1.82 ± 0.66	71.1 ± 22.6	63.4 ± 12.5
AG	1.86 ± 0.69	73.7 ± 23.4	64.7 ± 13.1
GG	1.93 ± 0.80	74.1 ± 22.3	63.7 ± 13.8
p-value	0.61	0.51	0.60
<i>GYPA</i> (rs2202507)			
AA	1.84 ± 0.66	72.8 ± 22.2	62.9 ± 11.8
AC	1.83 ± 0.64	72.5 ± 22.5	65.1 ± 12.8
CC	1.96 ± 0.80	73.9 ± 23.2	63.2 ± 13.4
p-value	0.33	0.89	0.27

Data are presented as mean ± SD, unless otherwise stated. SNP: single-nucleotide polymorphism; FEV₁: forced expiratory volume in 1 s; % pred: % predicted; FVC: forced vital capacity.

uncharacterised gene. The exact function or reason why this particular sequence is transcribed is unknown. Furthermore, the effect of the chromosome 4q31 locus (assigned through the rs2202507 SNP) has previously been attributed to the *GYPA* gene, despite a large physical distance separating the two (>200 kb). Glycophorins are a class of siaglycoproteins that are commonly found on the cell membranes of erythrocytes. The composition of the erythrocyte membranes are highly indicative of stress, as many different proteins found on the membrane are subject to oxidation and also change in expression in response to environmental stimuli, such as exposure to cigarette smoke. It has been shown that glycoporphins A, B and C show decreased expression in the erythrocytes of COPD patients, and can be used as a biomarker of oxidative stress [50]. Although the composition of *GYPA* on the membrane of red blood cells is indicative of oxidative stress, it is unclear how an alteration of the gene or its expression could influence susceptibility to either lung cancer or COPD. Whether the HHIP and *GYPA* associations with COPD and lung cancer reflect distinct effects [24] or an overlapping effect (through synthetic association [51]) at the 4q31 locus remains unknown.

There is good evidence to suggest that lung cancer and COPD are closely related, in that they share similar pathogenic pathways, *i.e.* inflammation, tissue remodelling and oxidant load. As such, they should not be considered in isolation but, rather, as two interrelated diseases [5–7] with overlapping genetic effects [30]. Approximately 10–15% of smokers will be diagnosed with lung cancer. Within this group, 50–90% of them will have pre-existing COPD. We and others have shown previously that smokers with reduced airflow, namely FEV1 consistent with COPD (GOLD stage II or higher), carry a four- to six-fold greater risk of lung cancer compared with smokers with normal lung function [5–7]. We conclude that FEV1 is a very good marker for susceptibility to lung cancer [7, 52]. Therefore, unless such phenotypes (COPD or reduced FEV1) are clearly defined in lung cancer studies, it is possible that the genetic associations reported to be linked to lung cancer may also (or actually) be associated with COPD [30]. Given the close relationship of these two diseases, it is not surprising that a genetic association with COPD may also be relevant to lung cancer. However, as spirometry is not routinely part of lung cancer studies [2, 20–22], the high prevalence of COPD in the lung cancer population may be missed. Similarly, a high prevalence of (undiagnosed) COPD may exist in smoking controls, obscuring the contribution of COPD-related SNP effects. In this study, we have used spirometry to carefully phenotype smokers in all cohorts and show that the 4q31 locus is equally important to both COPD and lung cancer. Due to the design of the current study, we can conclude that this relationship with lung cancer is not through a confounding or interactive effect with COPD [30]. The associations were unchanged after lung cancer cases were subdivided by the presence or absence of COPD. In future, replication of these results in larger populations with well-matched cohorts would strengthen these findings.

There are two relevant implications that stem from the finding that COPD and lung cancer share pathogenic pathways and an overlapping genetic susceptibility. First, predictive risk models for lung cancer should include COPD as a risk variable and/or

incorporate these overlapping genetic markers [13, 14]. Second, chemopreventive therapy targeting pathogenic pathways common to both COPD and lung cancer might confer considerable benefit to smokers and/or ex-smokers. With this in mind, it is interesting that guanosine triphosphatase inhibition by drugs such as 3-hydroxy-3-methyl-glutaryl-coenzyme A reductase inhibitors (statins) has been shown to inhibit lung matrix remodelling, reduce FEV1 decline, reverse EMT and be associated with lower lung cancer incidence [53]. We conclude that COPD and lung cancer are not solely interrelated through smoking or smoking-induced pathological processes in the lung: they are also linked through shared genetic effects conferred from overlapping loci, such as the chromosome 15q25 [30] and 4q31 loci.

STATEMENT OF INTEREST

A statement of interest for R.P. Young can be found at www.erj.ersjournals.com/site/misc/statements.xhtml.

REFERENCES

- Molfino NA. Genetics of COPD. *Chest* 2004; 125: 1929–1940.
- Broderick P, Wang Y, Vijayarishnan V, *et al.* Deciphering the impact of common genetic variation on lung cancer risk: a genome-wide association study. *Cancer Res* 2009; 69: 6633–6641.
- Kohansal R, Martinez-Camblor P, Agusti A, *et al.* The natural history of chronic airflow obstruction revisited: an analysis of the Framingham Offspring Cohort. *Am J Respir Crit Care Med* 2009; 180: 3–10.
- Mattson ME, Pollack ES, Cullen JW. What are the odds that smoking will kill you? *Am J Pub Health* 1987; 77: 425–431.
- Young RP, Hopkins RJ, Christmas T, *et al.* COPD prevalence is increased in lung cancer independence of age, sex and smoking history. *Eur Respir J* 2009; 34: 380–386.
- 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.
- Tockman MS, Anthonisen NR, Wright EC, *et al.* Airways obstruction and the risk for lung cancer. *Ann Intern Med* 1987; 106: 512–518.
- Wilson DO, Weissfeld JL, Balkan A, *et al.* Association of radiographic emphysema and airflow obstruction with lung cancer. *Am J Respir Crit Care Med* 2008; 178: 738–744.
- de Torres J, Bastarrika G, Wisnivesky JP, *et al.* Assessing the relationship between lung cancer risk and emphysema detected on low dose CT of the chest. *Chest* 2007; 132: 1932–1938.
- Mortensen EM, Coleland LA, Pugh MJ, *et al.* Diagnosis of pulmonary malignancy after hospitalisation for pneumonia. *Am J Med* 2010; 123: 66–71.
- Redline S, Tishler PV, Rosner B, *et al.* Genotypic and phenotypic similarities in pulmonary function among family members of adult monozygotic and dizygotic twins. *Am J Epidemiol* 1989; 129: 827–836.
- 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.
- Young RP, Hopkins RJ, Hay BA, *et al.* Lung cancer susceptibility model based on age, family history and genetic variants. *PLoS ONE* 2009; 4: e5302.
- Young RP, Hopkins RJ, Hay BA, *et al.* A gene based risk score for lung cancer susceptibility in smokers and ex-smokers. *Postgrad Med J* 2009; 85: 515–524.
- Schwartz AG, Ruckdeschel JC. Familial lung cancer: genetic susceptibility and relationship to chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2006; 173: 16–22.

- 16 Dasari V, Gallup M, Lemjabbar H, *et al.* Epithelial–mesenchymal transition in lung cancer: is tobacco the “smoking gun”? *Am J Respir Cell Mol Biol* 2006; 35: 3–9.
- 17 Shintani Y, Maeda M, Chaika N, *et al.* Collagen I promotes epithelial-to-mesenchymal transition in lung cancer cells *via* transforming growth factor- β signaling. *Am J Respir Cell Mol Biol* 2008; 38: 95–104.
- 18 Lee G, Walser TC, Dubinett SM. Chronic inflammation, chronic obstructive pulmonary disease, and lung cancer. *Curr Opin Pulm Med* 2009; 15: 303–307.
- 19 Sohal SS, Reid D, Soltani A, *et al.* Smoking has potential to initiate basement membrane disruption and epithelial mesenchymal transition in COPD. *Eur Respir J* 2009; 34: Suppl. 53, 2645.
- 20 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 Genet* 2008; 40: 616–622.
- 21 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.
- 22 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.
- 23 Pillai SG, Ge D, Zhu G, *et al.* A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. *PLoS Genet* 2009; 5: e1000421.
- 24 Wilk JB, Chen T, Gottlieb DJ, *et al.* A genome-wide association study of pulmonary function measures in the Framingham Heart Study. *PLoS Genet* 2009; 5: e1000429.
- 25 Repapi E, Sayers I, Wain LV, *et al.* Genome-wide association study identifies five loci associated with lung function. *Nat Genet* 2009; 42: 36–44.
- 26 Hancock DB, Eijgelsheim M, Wilk JB, *et al.* Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nat Genet* 2009; 42: 45–52.
- 27 Wilk J, Walter R, Laramie J, *et al.* Framingham heart study genome-wide association: results for pulmonary function measures. *BMC Med Genet* 2007; 8: Suppl. 1, S8.
- 28 Frayling TM. Genome-wide association studies provide new insights into type 2 diabetes aetiology. *Nat Rev Genet* 2007; 8: 657–662.
- 29 Stav D, Raz M. Prevalence of chronic obstructive pulmonary disease among smokers aged 45 and up in Israel. *Isr Med Assoc J* 2007; 9: 800–802.
- 30 Young RP, Hopkins RJ, Hay BA, *et al.* Lung cancer gene associated with COPD: triple whammy or possible confounding effect? *Eur Respir J* 2008; 32: 1–7.
- 31 Lambrechts D, Buysschaert I, Zanen P, *et al.* The 15q24/25 susceptibility variant for lung cancer and chronic obstructive pulmonary disease is associated with emphysema. *Am J Respir Crit Care Med* 2010; 181: 486–493.
- 32 Yang P, Yafei L, Jiang R, *et al.* A rigorous and comprehensive validation: common genetic variations and lung cancer. *Cancer Epidemiol Bio Prev* 2010; 19: 240–244.
- 33 Pearson TA, Manolio TA. How to interpret a genome-wide association study. *JAMA* 2008; 299: 1335–1344.
- 34 van Durme Y, Eijgelsheim M, Joos G, *et al.* The hedgehog interacting protein is a COPD susceptibility gene: the Rotterdam Study. *Eur Respir J* 2009; 36: 89–95.
- 35 Piitulainen E, Eriksson S. Decline in FEV₁ related to smoking status in individuals with severe α_1 -antitrypsin deficiency (PiZZ). *Eur Respir J* 1999; 13: 247–251.
- 36 Shi W, Chen F, Cardoso WV. Mechanisms of lung development: contribution to adult lung disease and relevance to chronic obstructive pulmonary disease. *Proc Am Thorac Soc* 2009; 6: 558–563.
- 37 Chuang PT, McMahon AP. Vertebrate Hedgehog signalling modulated by induction of a Hedgehog-binding protein. *Nature* 1999; 397: 617–621.
- 38 Cardoso WV, Lü J. Regulation of early lung morphogenesis: questions, facts and controversies. *Development* 2006; 133: 1611–1624.
- 39 Chuang PT, Kawcak T, McMahon AP. Feedback control of mammalian Hedgehog signaling by the Hedgehog-binding protein, Hip1, modulates Fgf signaling during branching morphogenesis of the lung. *Genes Dev* 2003; 17: 342–347.
- 40 Watkins DN, Berman DM, Burkholder SG, *et al.* Hedgehog signalling within airway epithelial progenitors and in small-cell lung cancer. *Nature* 2003; 422: 313–317.
- 41 Lemjabbar-Alaoui H, Dasari V, Sidhu SS, *et al.* Wnt and hedgehog are critical mediators of cigarette smoke-induced lung cancer. *PLoS ONE* 2006; 1: e93.
- 42 Fletcher C, Peto R. The natural history of chronic airflow obstruction. *Br Med J* 1977; 1: 1645–1648.
- 43 Løkke A, Lange P, Scharling H, *et al.* Developing COPD: a 25 year follow up study of the general population. *Thorax* 2006; 61: 935–939.
- 44 Moskvina V, Holmans P, Schmidt KM, *et al.* Design of case-controls studies with unscreened controls. *Ann Hum Genet* 2005; 68: 566–576.
- 45 Pritchard J, Stephens M, Donnelly P. Inference of population structure from multi-locus genotype data. *Genetics* 2000; 155: 945–959.
- 46 Thankkinstian A, Thompson JR, Minelli C, *et al.* Choosing between per-genotype, per-allele, and trend approaches for initial detection of gene-disease association. *J App Stat* 2009; 36: 633–646.
- 47 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.
- 48 Amos CI. Successful design and conduct of genome-wide association studies. *Hum Mol Genet* 2007; 16: 220–225.
- 49 Vestergaard J, Pedersen MW, Pedersen N, *et al.* Hedgehog signalling in small-cell lung cancer: frequent *in vivo* but a rare event *in vitro*. *Lung Cancer* 2006; 52: 281–290.
- 50 Minetti M, Leto TL, Malorni W. Radical generation and alterations of erythrocyte integrity as bio-indicators of diagnostic or prognostic value in COPD? *Antioxid Redox Signal* 2008; 10: 829–839.
- 51 Dickson SP, Wang K, Krantz I, *et al.* Rare variants create synthetic genome-wide associations. *PLoS Biol* 2010; 8: e1000294.
- 52 Young RP, Hopkins RJ, Eaton TE. Forced expiratory volume in one second: not just a lung function test but a marker of premature death from all causes. *Eur Respir J* 2007; 30: 616–622.
- 53 Young RP, Hopkins R, Eaton TE. Pharmacological actions of statins: potential utility in COPD. *Eur Respir Rev* 2009; 18: 222–232.