Genetic Predisposition to Chronic Obstructive Pulmonary Disease and/or Lung Cancer: Important Considerations When Evaluating Risk

Randa A. El-Zein1, Robert P. Young2, Raewyn J. Hopkins2, and Carol J. Etzel1

Abstract

Chronic obstructive pulmonary disease (COPD) is defined as a disease causing an airflow limitation that is not fully reversible. COPD is phenotypically complex and characterized by small-airway disease and/or emphysema that result from the interaction between host genetic susceptibility and environmental exposures. As in lung cancer, smoking exposure is the most important risk factor for the development of COPD, accounting for 80% to 90% of all cases. COPD affects an estimated 8% to 10% of the general adult population, 15% to 20% of the smoking population, and 50% to 80% of lung cancer patients (with substantial smoking histories). In prospective studies, COPD has been found to be an independent risk factor for lung cancer, conferring a three- to 10-fold increased risk of lung cancer when compared with smokers without COPD. These findings suggest that smokers have a host susceptibility to COPD alone, COPD and lung cancer (i.e., overlap), and lung cancer in the absence of COPD. This minireview focuses on important points that need to be addressed when studying genetic susceptibility factors for COPD and its complex relationship with susceptibility to lung cancer. Cancer Prev Res; 5(4); 522–7. ©2012 AACR.

Introduction

Chronic obstructive pulmonary disease (COPD) is an important public health problem, with approximately 24 million affected individuals and more than 125,000 annual deaths in the United States. The World Health Organization predicts that by 2020, COPD will become the third leading cause of mortality and the fifth leading cause of death worldwide (1). COPD has been found to be an independent risk factor for lung cancer, conferring a 3- to 10-fold increased risk of lung cancer when compared with smokers without COPD. Three separate absolute risk models for lung cancer have been developed: The Bach (2), Spitz (3), and Liverpool Lung Project (LLP; ref. 4) models. Well-documented risk factors such as smoking duration and occupational exposure to asbestos are common to all 3 models. The original Bach model did not include COPD. The LLP and Spitz models include lung-related comorbidities of COPD (Spitz) and pneumonia (LLP; ref. 4) models. Well-documented risk factors such as smoking duration and occupational exposure to asbestos are common to all 3 models. The original Bach model did not include COPD. The LLP and Spitz models include lung-related comorbidities of COPD (Spitz) and pneumonia (LLP; ref. 4) models.

Bach model (5). Recently, Maisonneuve and colleagues (6) updated the Bach model on the basis of data from the COSMOS screening trial, which included COPD. The updated Bach model showed a high discriminatory power and good calibration, which measures the agreement between observed results and the predictions of the model (6). In addition, Tammemagi and colleagues generated an extended Prostate, Lung, Colorectal, and Ovarian (PLCO) lung cancer risk model (7), which included COPD; this model has the highest discriminatory power reported to date. All of these models were based on predominantly Caucasian participants. In 2008, Etzel and colleagues (8) observed that COPD is also a major risk factor for lung cancer among African Americans and added it to the Spitz lung cancer risk model.

Cigarette smoking is the major cause of COPD, as of lung cancer, although only 15% to 30% of smokers develop either disease. Between 50% to 80% of lung cancer patients have preexisting COPD, compared with a 15% to 20% prevalence of COPD in the general smoking population (9–11), and studies consistently have shown that smokers who have COPD are at an increased risk for developing lung cancer (1, 12). Smokers with mild and moderate COPD have a 3-fold risk of developing lung cancer within 10 years, which increases to a 10-fold risk with severe COPD, when compared with smokers with normal lung function (as determined by the specific spirometric criteria FEV1/FVC less than 70 and FEV1% predicted less than 80%; ref. 1). It has been reported that the risk of developing lung cancer does not disappear completely after smoking cessation, with a roughly equal proportion of lung cancers being...
diagnosed in former smokers or current smokers (approximately 40% to 45% each; refs. 9, 13). The damage induced by cigarette smoke in COPD patients is not fully reversible and also persists in these patients after smoking cessation, thus explaining the persistence or even the progression of the disease in former smokers (14).

COPD biology

COPD is a phenotypically complex disease characterized by small-airway disease and/or emphysema that results from the interaction between host genetic susceptibility and environmental exposures; this disease is not fully reversible. The major cause of death among COPD patients is lung cancer. Given the same degree of pulmonary function impairment and smoking history, patients showing predominantly small-airway disease are at a greater risk of lung cancer compared with those with a predominance of emphysema (15). For patients with predominantly emphysema who do develop lung cancer, however, the survival rate following surgery for early-stage lung cancer is lower than for lung cancer patients with airway COPD (16). It has been suggested that the mixed-phenotype disease lung cancer includes COPD defined as a subphenotype by the pulmonary function test spirometry (9, 11, 17). The mechanisms linking COPD to lung cancer remain unclear; a number of hypotheses are being tested, however, such as a shared genetic risk between these 2 disorders (17) and delayed clearance of inhaled carcinogens in COPD patients. Cigarette smoke contains a number of known carcinogens and a very high concentration of oxidants (18) that in combination induce inflammation leading to DNA damage and mutations in lung and airway tissues. With airflow impairment, patients are not able to fully clear the tobacco carcinogens, thus increasing the opportunity to induce DNA damage and mutations (19). In addition, the chronic and persistent oxidative stress and local inflammation have also been implicated in the pathogenesis of lung cancer (20). To date it is unclear how COPD contributes to lung cancer risk or whether both COPD and lung cancer are the result of common underlying exposures or whether some combination of both scenarios applies.

COPD is a genetically complex disease in which about 1% of affected patients are α-1 antitrypsin deficient and hence genetically predisposed to the development of the disease. This deficiency accounts for a small proportion of COPD among never smokers. The majority of COPD cases reflect a complex interplay between genetic and environmental interactions. Several studies have reported on the association between genetic variants and COPD, with inconsistent results. The discrepancy in reported findings could be attributable to several issues such as improper selection of the candidate genes, study sample size (21), the unclear definitions or smoking habits of cases and controls, and a lack of proper disease classification in lieu of simply confirming the presence or absence of disease (17).

Genome-wide association studies (GWAS) of lung cancer and COPD have independently identified several chromosomal regions and candidate genes, including chromosome 1q21 (CRP, IL-6R), 4q22, -24, and -31 (FAM13A, GSTD1, HHIP, and GYPA), 5p15 (CRBR9), 5q32 (HTF4, ADAM19), 6p21, 6q24 (BAT3, AGER1, and GPR126), 15q25 (CHRNA3/5), and 19q13 (RAB4B, EGLN2, MIA, and CYP2A6; refs. 22–32), that are associated with host susceptibility to the development of lung cancer and/or COPD, with some loci showing significant overlap between COPD and lung cancer. Young and colleagues reported a case-control study in which all subjects were phenotyped for COPD according to spirometry; the CHRNA3/5 (15q25) and HHIP (4q31) loci were associated with both COPD (heritability estimated to be 40%–75%) and lung cancer (heritability estimated to be 15%–25%, an estimate based on concordance in twin studies adjusted for smoking; refs. 33, 34). A variant in the FAM13A gene on 4q22 (associated with lung function; ref. 25) is similarly linked to a reduced risk of COPD and lung cancer (35). It is interesting that extending the lung cancer risk models by addition of the top single-nucleotide polymorphism (SNP; P < 10⁻⁵) identified through GWAS has not necessarily further improved the discriminatory power of models compared with their epidemiology-based counterparts because these top SNPs confer only small to modest degrees of risk for disease, even though each individual SNP has genome-wide significance (36). Likewise, extension of the PLCO lung cancer risk model by adding pulmonary function or other markers such as sputum DNA image cytometry (SDIC) did not substantially improve the discriminatory power of the model, although both pulmonary function and SDIC are significantly associated with lung cancer risk (37). As Tammemagi and colleagues explained, factors which are significantly associated with lung cancer may not contribute substantially to prediction because association is measured with effect estimates and prediction is measured by estimating discrimination and calibration (37).

Several recent studies have focused on the role of genetic variants in association with smoking behavior. In more than 10,000 participants from the general population of Denmark, Kaur-Knudsen and colleagues (38) examined the association between nicotinic acetylcholine receptor genotype and smoking behavior and the added effect of this association on the risk of tobacco-related diseases such as lung and bladder cancer, COPD, and cardiovascular diseases. They reported an association between the nicotinic acetylcholine receptor genotype and daily tobacco consumption, cumulative tobacco consumption, and smoking inhalation but not with age of smoking onset, age of smoking cessation, or smoking duration. Furthermore, a nicotinic acetylcholine receptor polymorphism was associated with an additional increased risk of lung and bladder cancer and COPD after adjustment for smoking. Hamidovic and colleagues (39) investigated the association between genetic variability and smoking persistence in African Americans and identified a locus downstream of the brain-derived neurotrophic factor 3′-untranslated region as a mediator of smoking behavior. In addition, smoking persistence in African Americans was associated with independent variants in the cluster of genes encoding nicotinic...
Considerations when attempting to identify susceptibility factors for COPD

Wang and colleagues (49) recently reported that the CHRNA5-A3 region on chromosome 15q24–25 is not only associated with lung cancer risk but also with lung cancer risk through its effects on both smoking exposure and COPD. Using mediation analysis methods, the authors concluded that COPD is a mediating phenotype that explains part of the effect of smoking exposure on lung cancer and that smoking behavior is a mediator of the relationship between the SNP rs1051730 (from the chromosome 15q25 locus) and COPD risk. Previously, Spitz and colleagues and et al (3, 8) included COPD as an independent risk factor for lung cancer in their risk assessment models among Caucasians and African Americans. Therefore, (i) including COPD as an independent or mediating variable in risk models evaluating genetic susceptibility to lung cancer is crucial for an accurate risk assessment for lung cancer and (ii) participants with and without COPD are needed in evaluations of the susceptibility to lung cancer. When elucidating susceptibility factors for COPD among smokers, however, several issues (as follow) need to be considered.

Choice of controls

It is well documented that the prevalence of COPD in smokers enrolled as study controls in any epidemiologic study is highly dependent on the recruitment method and sampled population. The prevalence of COPD is 30% to 50% in “convenience samples” such as community-based smoking volunteers, computed tomography (CT) screening participants, or random hospital/clinic–based controls. The true prevalence of COPD, as shown by spirometry evaluation in randomly selected populations of smokers, is about 10% to 20%. Spirometry is the internationally accepted method of diagnosing and classifying COPD. Thus in the absence of spirometry findings, the prevalence of COPD in lung cancer cases and control smokers is likely to be different; based on the discrepant prevalences in convenience sample versus randomly selected smokers it could be argued that smokers matched for ethnicity, gender, and pack-years and with normal or near-normal lung function would minimize any potential to misclassify smokers with COPD as “controls” (50–53). Matching (or stratifying) for smoking exposure is particularly important because the clinical expression of genes conferring “susceptibility” or “resistance” to COPD (i.e., penetrance) is dependent on the smoking exposure dose (cigarettes per day, years smoked, and total pack-year exposure). This dependence makes COPD (and lung cancer) “dynamic” phenotypes, occurring almost exclusively after exposure to smoking (or other aero-pollutants; ref. 17). Similarly, healthy or resistant smokers (controls) for COPD studies can only be reliably identified by spirometry. Comparing COPD smokers with healthy/resistant control smokers is critical for identifying protective genetic effects (34, 35).

We note that the common approach of adjusting for variation in smoking exposure between cases and controls (rather than stratifying) may be problematic if there is a threshold effect on COPD (48, 54) or if the relationship of smoking exposure dose to lung function is not linear (55). Conversely, it is as important to exclude nonsmokers from the case and control comparison groups because including nonsmokers may obscure or attenuate any genetic effects (especially with a disproportion of nonsmokers in cases compared with controls), as they have an insufficient exposure to tobacco smoke carcinogens. Recent studies suggest that the etiologies of COPD among never smokers and ever smokers are potentially quite different, as exemplified by the 10% to 15% of COPD attributable to α1-antitrypsin deficiency, dust exposure, or chronic asthma. Furthermore, Li and colleagues and Sampsonas and colleagues reported that different genes may be relevant in nonsmokers (vs. smokers) with lung cancer or COPD (56, 57).

Classification of COPD patients

The accurate classification of the clinical phenotypes of COPD is essential to the success of identifying genetic variants associated with such a complex and phenotypically variable disease. Such classification would refine the current concept of COPD from a unique disease to a syndrome with multiple phenotypes that are expressed because of different underlying pathobiologic processes (58–61). To date, the most important and most studied phenotype of COPD is airflow limitation defined by pulmonary function testing (spirometry). In 2010, however, the International COPD Genetics Conference reported that additional phenotypic expressions should be considered (21). These phenotypes should include the degree, type, and distribution of emphysema; extent of thickening of airway walls; presence and degree of dyspnea; quality of gas exchange in the lungs; and presence of systemic inflammation. The report concluded that determination of these phenotypes will not only provide interesting scientific mechanisms for the disease but also will have clinical prognostic and therapeutic value. It should be stressed that although a medical history of COPD correctly identifies COPD in 80% to 90% of cases (a small proportion have asthma), history alone is not sufficiently accurate to classify (or phenotype) all subjects in genetic epidemiologic studies. When depending solely on diagnosis through symptoms and medical history, between 50% to 70% of COPD is missed because of a lack of lung function testing (spirometry; refs. 49–51).
especially in asymptomatic smoking controls or lung cancer patients (in whom routine spirometry suggests that 50%–70% have COPD).

**COPD–lung cancer "overlap" and selection of cases and controls**

Whether COPD and lung cancer are related directly through shared genetic susceptibility or through mechanisms related to the differential effect of smoking exposure and/or inflammation in smokers with COPD, it is important to subphenotype for COPD in any epidemiologic study of lung cancer. The evidence to date suggests that smokers who develop COPD—either COPD and then lung cancer or lung cancer in the absence of COPD—are likely to be distinct phenotypic groups. This subphenotyping is not as easy as it sounds because age is also relevant to the correct phenotyping of smokers; the vast majority of COPD and lung cancer cases are diagnosed in people 50 years or older (including up to 90 years old). To date, the vast majority of genetic studies of COPD and lung cancer have been done independently of each other, with no regard for the mediating or confounding effect of COPD on lung cancer (or vice versa). Therefore, any genetic associations reported for lung cancer may in fact be related to or mediated by COPD, a finding described by Young and colleagues in several published studies (22, 33, 62). de Andrade and colleagues (63) recently used a convenience sample of lung cancer cases with and without COPD to evaluate the effect of genetic variants in glutathione metabolism, DNA repair, and inflammatory response pathways to identify susceptibility markers for COPD risk. It is particularly problematic to use samples of convenience in genetic epidemiologic studies to identify susceptibility or protective genetic variants because patients with COPD and lung cancer may not represent those with COPD alone.

One way to overcome this issue is to subphenotype lung cancer cases by pulmonary function tests and chest CT into those with and without COPD and then test for associations separately (64). This ascertainment might be possible in the COPDGene study currently underway in the United States (www.COPDGene.org). It may also be possible in CT screening trials, in which both pulmonary function tests and baseline CT data are available on smokers who are followed prospectively. In the Pittsburgh Lung Screening Trial (10), nearly 50% of screenees had COPD (on the basis of pulmonary function tests and/or CT evidence of emphysema at baseline), compared with COPD in 85% of those who subsequently developed lung cancer (Fig. 1). This study suggests that 85% of lung cancers detected prospectively in a CT screening trial had underlying evidence of a disposition to COPD at baseline and that this disposition accounted for about 5-fold more lung cancers (detection rate of 5%) than did normal lungs at baseline (detection rate of 1%; ref. 10 and Fig 1). It has been suggested that a COPD-based and/or gene-based approach to CT screening for lung cancer may help better target people at the greatest risk (11, 17, 47, 64, 65). These approaches are but examples of the potential clinical use to which a better understanding of the relationship between COPD and lung cancer and, in particular, of the contribution of genetic susceptibility (or resistance) to adverse smoking outcomes can be put.

**Conclusion**

Epidemiologic studies to date indicate that COPD, defined by spirometric criteria, is common among smokers (estimated to affect about 20%) and even more common among smokers diagnosed with lung cancer (estimated to affect about 50%–70%). Although the basis of this strong association is not yet understood at a molecular level, genetic epidemiologic studies are now uniquely placed to clarify this important relationship. However, future studies must consider COPD–lung cancer overlap by carefully phenotyping all study participants according to accepted criteria. Studies of the genetic predisposition to COPD and lung cancer should avoid samples of convenience and should take into consideration (i) the importance of smoking exposure and age-to-gene penetrance and (ii) the complex subphenotypes underlying smoking-related lung disease such as small-airway disease and emphysema. Ideally, these prospective studies should incorporate data on pulmonary function testing and CT-detected emphysema into the design.

**Disclosure of Potential Conflicts of Interest**

RPY has been supported by funding from Synergenz Biosciences Ltd., which holds patents for gene-based risk testing for COPD and lung cancer susceptibility.
Cancer Prevention Research

Authors' Contributions

Conception and design: R. El-Zein, C.J. Etzel
Development of methodology: R. El-Zein, C.J. Etzel
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): R. El-Zein, C.J. Etzel
Analysis and interpretation of data (e.g., statistical analysis, biosstatistics, and computational analysis): C.J. Etzel
Writing, review, and/or revision of the manuscript: R. El-Zein, R. P. Young, R.I. Hopkins, C.J. Etzel

Administrative, technical, or material support (i.e., reporting or organizing, data, constructing databases): R. El-Zein, C.J. Etzel

Grant Support

R.P. Young and the funding of his research has been supported by grants from the University of Auckland, Health Research Council of New Zealand and Synergygen Biosciences Ltd.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to identify this fact.

Received December 7, 2012; revised January 11, 2012; accepted January 26, 2012; published online April 4, 2012.

References
