Molecular epidemiology of lung cancer

P. G. Shields
Molecular Epidemiology Section, Laboratory of Human Carcinogenesis, Division of Basic Sciences, National Cancer Institute, Bethesda, MD, USA

Summary
Lung cancer occurs through a complex multistage process that results from the combination of carcinogen exposure and genetic susceptibilities. The primary etiology of lung cancer is tobacco smoking, but an understanding of why some smokers develop lung cancer, and others do not, remains unclear. Current studies focus on genetic susceptibilities to lung cancer, and how they modify the effects of tobacco smoke carcinogens. New assays are being developed to study other contributors to cancer risk, such as interindividual differences in DNA repair. There is current evidence to suggest that the risk of lung cancer for women, compared to men, is higher for the same level of smoking. Several biological differences for the types of lung cancer have been observed in women and men. Also, there appear to be differences in lung cancer between Caucasians and African-Americans. Molecular epidemiology tools are uniquely suited to study these biological differences. These studies will improve cancer risk assessments and focus cancer prevention strategies. Other studies also are focusing on tobacco addiction, in order to lead to improved smoking cessation strategies.

Key words: carcinogen metabolism, genetic susceptibility, risk assessment

Introduction
According to the American Cancer Society Facts and Figures [1], lung cancer is the leading cause of cancer-related death in men and women. The predominant cause of lung cancer is tobacco smoking; more than 90% of lung cancer patients are smokers [2]. Tobacco smoke contains more than 100 diverse mutagens and carcinogens. This complex exposure leads to multiple genetic defects, rapid cell proliferation and resistance to therapy. The suspected carcinogens in tobacco smoke include, but are not limited to, polycyclic aromatic hydrocarbons, N-nitrosamines (including tobacco specific nitrosamines) and maybe aromatic amines. The metabolites of these carcinogens are direct acting mutagens, although they also might cause DNA damage through oxidative damage. In spite of their enormous exposure, only about 10% of heavy smokers suffer from lung cancer. The reasons for this are being explored using molecular epidemiological tools.

The field of molecular epidemiology seeks to identify human cancer risk based upon individual exposures and susceptibilities to cancer. The overall goal is to develop, apply and validate biomarkers of human cancer risk, in order to enhance cancer risk assessments, focus cancer prevention strategies and elucidate mechanisms of carcinogenesis. Two fundamental principles of molecular epidemiology relate to the complexity of carcinogenesis and the role of interindividual variation in host capacity for carcinogenic responses. For the first, studies are designed based upon the knowledge that cancer is a multistage process where each stage is composed of several genetic events, and each event occurs because of multiple steps requiring carcinogen metabolic activation, detoxification, mutations in critical parts of cancer-related genes, and failure to repair DNA damage or undergo programmed cell death. The second fundamental principle of molecular epidemiology is based upon data demonstrating interindividual variation [3] in carcinogen metabolism and associated responses.

Gene-environment interactions for lung cancer
Numerous gene-environment interactions for lung cancer risk have been demonstrated, whereby genetic polymorphisms in carcinogen metabolizing genes have been shown to modify the effects of carcinogen exposure on cancer risk and DNA adducts [4–7]. Those genes associated with lung cancer risk include the glutathione-S-transferase M1 (GSTM1), cytochrome P450 (CYP) 2D6, CYP1A1, CYP2E1, and N-acetyltransferase 2 (NAT2). There are certainly other genes that are plausibly related to lung cancer risk including GST-Pi, CYP2A6, CYP1B1, CYP3A4, microsomal epoxide hydrolase (MEH), Naceytransferase 1 (NAT1), aryl hydrocarbon hydroxylase receptor, and others, although not all of these have been shown to have functional genetic polymorphisms.
There are several studies of the CYP1A1 gene, especially in Japanese subjects, that indicate its importance in cancer risk. Three genetic polymorphisms related to cytochrome p450 (CYP1A1) have been identified [8-10, 12], two of which are possible markers of lung cancer risk in the Japanese [9, 11] and one of which exists only in African-Americans [10]. The first to be studied [9] was an RFLP located 3' to the last exon of the gene and revealed by enzymatic digestion with the Mspl restriction enzyme. An increased quantitative change in the protein product resulting from this polymorphism has been reported in one [12], but not another study [13]. The second polymorphism is an A: T → G: C base substitution responsible for an isoleucine to valine substitution (residue 462) in the catalytic heme binding region [8]. In the Japanese, the presence of the less common (minor), but more active allele at the Mspl site was predictive of the valine substitution [8, 14]. Further, it was found that there was an interaction between this polymorphism and smoking for lung cancer risk [11, 14], where the combination of the homozygous minor allele and smoking yielded odds ratios similar to having one of the other genotypes and a greater smoking history. There is also one report that associates this polymorphism with increased stages of disease [14], and recent data suggest that CYP1A1 genetic polymorphisms predict the presence of p53 mutations [15]. Nonetheless, an association of CYP1A1 polymorphisms and lung cancer risk has only been identified in Japanese. Several Western studies of the Mspl locus, including Caucasians [16-19] and African-Americans [16] did not find an increased risk. For the valine substitution, several studies [17, 19] in Northern Europe have also been negative, although one report from Brazil has found a positive association in persons with squamous cell lung cancer [20]. For the intron 3 polymorphism specific to African-Americans [10], the initial indication was that there was an association with lung cancer [10], but another study was negative [21].

A presence of a homozygous deletion of the glutathione-S-transferase gene (GSTM1) results in the loss of function [22, 23] and the ability to conjugate and detoxify carcinogens. The loss of detoxification ability has been associated with increased PAH-DNA adducts in human lung [49] and in bronchial epithelium in particular [50]. The enzyme is known to conjugate PAH epoxide intermediates. A recently described polymorphism in codon 104 of GST-Pi is associated with lung cancer risk and the presence of hydrophobic DNA adducts [51].

While most previous studies of inherited susceptibility have focused on carcinogen metabolism, certainly human variability in DNA repair and programmed cell death will also contribute to lung cancer risk. One method that can be used to phenotypically characterize persons for DNA repair has been named the 'mutagen sensitivity assay' [52]. Briefly, blood is collected from study subjects, their lymphocytes are cultured, followed by an exposure to a clastogen, and then chromosomal aberrations are counted by giemsa staining. For cancer risk in the general population, bleomycin is used as the clastogen. (Bleomycin is a chemotherapeutic agent that does not require metabolic activation.) Here, individuals can be categorized as sensitive or resistant, depending on the number of chromosomal breaks. An increased number of breaks, which identifies the 'sensitive individual' was associated with lung cancer risk in African-Americans [53], and was positive for both adenocarcinoma and squamous cell cancer. There also was an interaction between smoking and mutagen sensitivity for cancer risk, although the number of cases was small after stratification. Notably, there was no relation between smoking and sensitivity. A genetic polymorphism in GSTM3 has been reported [39], but not yet studied in lung cancer.

Gene-gene interactions for CYP1A1 and GSTM1 have been found in ongoing Japanese studies [40]. The combined 'at risk' genotypes resulted in an odds ratio of up to 41 for the highest levels of smoking. Also, while each 'at risk' genotype conferred a higher probability for presenting with more extensive cancers [14, 41], the combined genotypes also yielded a higher risk and shorter survival [41].

There are several current genetic polymorphisms that are suggested to be lung cancer risk factors, but have not yet been extensively studied. The CYP2A6 genetic polymorphism can be studied phenotypically by administering coumarin (CYP2A6 is responsible for coumarin hydroxylase activity) [42] or by recently described genetic-based methods [43]. The protein specifically activates N-nitrosamines, including tobacco specific nitrosamines (NNK) [44, 45], which results in mostly G→A transitions [46] and cell transformation [47, 48], using in vitro cell culture models. Thus, it both activates carcinogens and can be directly related to amount of exposure. The coumarin hydroxylase activity is associated with a polymorphism in the gene, where the CYP2A allele results in a single inactivating amino acid change and decreased activity [43]. The allelic frequency of this variant is 15% in Caucasians and 0% in African-Americans.

The GST-Pi is among the most abundant GSTs in the human lung [49] and in bronchial epithelium in particular [50]. The enzyme is known to conjugate PAH epoxide intermediates. A recently described polymorphism in codon 104 of GST-Pi is associated with lung cancer risk and the presence of hydrophobic DNA adducts [51].

A genetic polymorphism in GSTM3 has been reported [39], but not yet studied in lung cancer.

Gene-gene interactions for CYP1A1 and GSTM1 have been found in ongoing Japanese studies [40]. The combined 'at risk' genotypes resulted in an odds ratio of up to 41 for the highest levels of smoking. Also, while each 'at risk' genotype conferred a higher probability for presenting with more extensive cancers [14, 41], the combined genotypes also yielded a higher risk and shorter survival [41].

There are several current genetic polymorphisms that are suggested to be lung cancer risk factors, but have not yet been extensively studied. The CYP2A6 genetic polymorphism can be studied phenotypically by administering coumarin (CYP2A6 is responsible for coumarin hydroxylase activity) [42] or by recently described genetic-based methods [43]. The protein specifically activates N-nitrosamines, including tobacco specific nitrosamines (NNK) [44, 45], which results in mostly G→A transitions [46] and cell transformation [47, 48], using in vitro cell culture models. Thus, it both activates carcinogens and can be directly related to amount of exposure. The coumarin hydroxylase activity is associated with a polymorphism in the gene, where the CYP2A allele results in a single inactivating amino acid change and decreased activity [43]. The allelic frequency of this variant is 15% in Caucasians and 0% in African-Americans.

The GST-Pi is among the most abundant GSTs in the human lung [49] and in bronchial epithelium in particular [50]. The enzyme is known to conjugate PAH epoxide intermediates. A recently described polymorphism in codon 104 of GST-Pi is associated with lung cancer risk and the presence of hydrophobic DNA adducts [51].

While most previous studies of inherited susceptibility have focused on carcinogen metabolism, certainly human variability in DNA repair and programmed cell death will also contribute to lung cancer risk. One method that can be used to phenotypically characterize persons for DNA repair has been named the 'mutagen sensitivity assay' [52]. Briefly, blood is collected from study subjects, their lymphocytes are cultured, followed by an exposure to a clastogen, and then chromosomal aberrations are counted by giemsa staining. For cancer risk in the general population, bleomycin is used as the clastogen. (Bleomycin is a chemotherapeutic agent that does not require metabolic activation.) Here, individuals can be categorized as sensitive or resistant, depending on the number of chromosomal breaks. An increased number of breaks, which identifies the 'sensitive individual' was associated with lung cancer risk in African-Americans [53], and was positive for both adenocarcinoma and squamous cell cancer. There also was an interaction between smoking and mutagen sensitivity for cancer risk, although the number of cases was small after stratification. Notably, there was no relation between smoking and sensitivity.
Racial and gender differences for lung cancer

Although all races have an increased risk of lung cancer due to smoking, the risk of lung cancer by race appears to be different, even for the same level of exposure to tobacco smoke. For example, studies at the American Health Foundation [54] indicated that there was a relative risk of 1.8 (95% CI: 1.4–2.2) for African-Americans compared to Caucasians. Risks for African-Americans compared to Caucasians are higher at lower levels of smoke exposure. In a separate case-control study of different ethnicities [55], following adjustment for pack years, occupation, education and age, compared to the Japanese, there was a 121%, 53% and 46% difference in risk for Hawaiian, Filipino, and Caucasian males. However, there has been little biological evidence to support an increased risk for African Americans. Preliminary data indicates that there is an increased level of \( N \)-nitrosamines adducts in the lungs of African-American smokers compared to Caucasian smokers (unpublished data), as well as an increase in urinary metabolites for tobacco-specific nitrosamines [56].

Epidemiological evidence also indicates that women are at a relatively higher risk of lung cancer for a given degree of smoking. Within levels of smoking, a relative risk ranging from 1.7 to 2.9 has been reported [54, 57, 58] for women compared to men. The levels of risk did not appreciably differ by levels of smoking exposure. Corroboration laboratory data supports an increased risk for women. It has been found that DNA adducts in lung tissue of women are higher than in men using a non-specific \( 32\)P-postlabeling assay that tends to identify bulky adducts such as those from PAHs [7], and from a \( 32\)P-postlabeling assay for \( N \)-nitrosamines (unpublished data). Also, while men have more p53 mutations overall, the level of G6T transversions are higher in women, suggesting a greater biological impact for smoking. Separately, the incidence of lung cancer types, i.e., Kreyberg I versus Kreyberg II, is different for men and women, where the latter predominates for women [54, 59], and the difference is more pronounced for non-smoking women [60]. The reasons why women compared to men are more susceptible to tobacco are unknown. It is possible that risk increases due to circulating endogenous sex-steroids (estrogens and progesterone), or from estrogen use (oral contraceptives or estrogen replacement therapy [ERT]). The role of estrogens might relate to increased carcinogen metabolism through the induction of cytochrome \( p450 \)s, or by causing DNA damage from estrogen metabolites or oxidative damage. Sex steroids also can play a role in lung cancer by enhancing cell proliferation; it has been demonstrated that lung cancer cells contain both estrogen and progesterone receptors, but that they are present in 66% of women and 33% of men. The relation of receptors to risk is unknown but could be explored in this study setting.

Tobacco addiction

The amount of cigarette smoking, which is directly related to lung cancer risk, depends upon many factors including the need to maintain nicotine levels in the blood, satisfaction of the brain’s reward pathways, cost of cigarettes, and advertising. The need to maintain nicotine levels and satisfaction of reward pathways may be very strong, as persons who try to consume fewer cigarettes tend to inhale more deeply, thereby maintaining overall exposure to tar but smoking fewer cigarettes [61, 62]. Nicotine has a ‘rewarding’ property that serves to reinforce drug seeking behavior [63–65]. Central nicotinic acetylcholine receptors are stimulated by nicotine, and are upregulated and desensitized simultaneously by chronic exposure. These receptors stimulate the secretion of dopamine into the neuronal synapse, which then stimulates post-synaptic dopamine receptors, thereby satisfying craving. Host susceptibilities may play a role in nicotine dependence, specifically metabolizing enzymes governing nicotine levels in the body and neuro-behavioral factors relating to the reinforcing value of nicotine. The former might dictate the initial pharmacological reactions to nicotine and how much smoking is needed to maintain nicotine levels [63, 64], while the latter may affect why people need to maintain nicotine levels.

We have hypothesized that interindividual variation for dopamine pathways and the reward mechanism might lead to an increased risk of smoking. To examine this hypothesis, we have studied polymorphisms in genes that govern synaptic dopamine levels through active reuptake by the dopamine transporter and in dopamine receptors [66] and have recently shown that the risk of smoking is related to a genetic polymorphism in the dopamine reuptake transporter gene, and that there is an interaction with a dopamine D2 receptor polymorphism (Lerman et al., unpublished data). A genetic polymorphism in the dopamine D4 receptor was found to statistically interact with depression to increase smoking risk [67]. Also, the presence of other dopamine D4 receptor alleles has been associated with the risk of smoking in African-Americans, as well as the ability to quit smoking following therapy (Shields et al., unpublished data). Other genetic polymorphisms have been studied, such as those that relate to dopamine metabolism via the tyrosine hydroxylase gene, but no association was found [68].

Acknowledgements

The author would like to thank Drs C. Lerman and N. Caporaso for there exciting collaboration with the tobacco addiction studies, and Dr C. Harris for his numerous types of support.
References

40. Nakachi K, Imai K, Hayashi S, Kawajiri K. Polymorphisms of the CYP1A1 and glutathione S-transferase genes associated with...