Glutathione-S-Transferase polymorphisms as risk factors for lung cancer: A case-control study

Caralee E. Caplan

A. Study Purpose and Rationale

Lung cancer is the leading cause of cancer death in the United States, and despite advances in medical and surgical treatments, the five-year survival has remained stable at 10-14% over the past three decades. Although smoking cessation is an important public health goal, many smokers find it difficult or impossible to quit and even patients who have successfully abstained continue to represent 50% of newly diagnosed cases. Because the only long-term survivors of lung cancer are patients with early-stage tumors, recent efforts, among them the Early Lung Cancer Action Project (ELCAP), have focused on screening techniques to diagnose cancer at an earlier stage and improve survival. The ELCAP study found that low-dose computed tomography (CT) scanning in asymptomatic smokers is more sensitive than chest x-ray and is able to detect sub-centimeter nodules and early-stage lung cancers. (1) However, fewer than 10% of the nodules detected were associated with a definitive diagnosis of cancer; the vast majority of nodules of indeterminate clinical significance are benign, but some will be malignant. It is our hypothesis that concurrent examination of molecular epidemiological markers of cancer risk assessment may enhance the efficacy of low-dose CT screening for lung carcinoma by helping to identify individuals at highest risk for a lung neoplasm.

The search for such epidemiological markers was inspired by the observation that despite the clear association between cigarette smoking and lung cancer, only approximately 10-20% of smokers develop neoplasms of the lung. (2) Family history clearly plays a role. Investigators have therefore hypothesized that genetic susceptibility factors and other environmental exposures must therefore modulate the DNA damaging effect of smoking on airway epithelial cells. (3) Indeed, most experts in the growing field of molecular epidemiology believe that an individual's susceptibility to cigarette smoke-related lung cancer is, in part, determined by the balance between the capacity to activate inhaled procarcinogens to carcinogens (phase 1) and to detoxify carcinogens (phase 11). The prevailing model is that phase I metabolism is performed by several enzymes of the cytochrome P450 superfamily (e.g., CYP I A I) and that phase II metabolism is continued by other enzymes, most notably, the glutathione-S-transferases (GST). Once carcinogens are formed and the detoxifying mechanisms fail, DNA damage may occur, leading to disregulation of cellular proliferation and ultimately, to lung cancer. (2,3)

To date, numerous case-control studies have attempted to determine whether certain polymorphisms of carcinogen-metabolizing enzyme systems confer an increased risk of lung cancer. GSTs are a multigene family of enzymes that conjugate electrophilic carcinogens (such as polycyclic aromatic hydrocarbons, PAHs) with glutathione and render them more readily excreted. Within the human GST-mu class, three different variant alleles have been detected. While GSTMI *A and GSTMI *13 code for enzymes with similar activity, the GSTM 1* 0 (null allele) produces no active enzyme. The GSTMI null phenotype is a very common polymorphism, occurring in about 50% of Japanese, European, and North American control populations. (2,3) Because even a modestly elevated cancer risk associated with this highly prevalent phenotype would have significant public health implications, many investigators have examined the hypothesis that the GSTMI null phenotypes and genotypes are over-represented in lung cancer patients. A recent study found an OR of 2.6 (95% Cl 1.1 -6.1) for *GSTMI* deletion in female nonsmokers with environmental tobacco smoke exposure compared to female nonsmokers without such exposure. (4) Furthermore, a positive association between *GSTMI* deletion and PAH-DNA adducts (an indicator of DNA damage) has been reported for lung and mononuclear cells. (5,6) Another study investigating the relationship between *GSTMI* genotype and PAH-DNA in

mononuclear cells of smokers found the highest levels of adducts in subjects with the *GSTMI* null genotype, although the difference was not statistically significant, likely due to small sample size. (7)

Although several studies have found an association between the GSTM I null genotype and lung cancer, odds ratios tend to be low (less than 2), and many studies have found no association. In a recent meta-analysis of 23 case-control studies, the authors found an overall odds ratio of lung cancer risk associated with GSTMI deficiency was 1. 13 (95% Cl, 1.04-1.25) for studies employing genotyping methods. (8,9) However, the meta-analysis underlines some key flaws in the design of many of the case-control studies. Few studies are adequately powered to detect moderate effects, most fail to take important sources of bias-such as ethnicity, age, and gender-into account, and many fail to examine the role of combinations of genotypes, which are likely more discriminating as risk factors than a single locus. Furthermore, because lung tissue is relatively inaccessible, most studies have used blood leukocytes as a surrogate tissue, an approach which may or may not be entirely valid. (2)

The results relating to GSTTI, another sub-group of enzymes of the GST class, have been similarly variable. However, at least 2 studies have shown that the combined GSTM1 and GSTT I null genotype may confer a slightly increased risk of cancer in longterm smokers. (10, 11)

The purpose of this study is to re-examine the hypothesis that the frequency of the GSTMI null genotype and the GSTTI null genotype (separately and combined) will be higher in cases than in controls. We plan to account for other known risk factors for lung cancer and factors that may affect genotype distribution (e.g., age, gender, ethnicity, smoking history). We also plan to collect not only blood but also buccal cells and sputum samples as surrogate tissues.

B. Study Design and Statistical Analysis

This study will employ a case-control design to determine odds ratios of lung cancer risk associated with the GSTMI null genotype, the GSTTI null genotype, and both combined. Most of the case-control studies to date have suffered from one of the most important pitfalls of the case-control design-the difficulty in finding appropriate controls. In our study, this problem is virtually eliminated, because cases and controls will be drawn from the same population, and cases identified at a preclinical stage.

The large NY-ELCAP study intends to enroll approximately 10,000 patients from 8 centers in New York City and 2 centers elsewhere in New York State for CT screening for early detection of lung cancer. The subjects will be men and women older than 60 years of age with a smoking history of 10 pack-years or greater. Based on the initial ELCAP data, which identified 27 malignant nodules among 1000 subjects at baseline screening, yielding a prevalence rate of about 2.7%, we would expect to identify approximately 270 cancers at baseline screening of the 10,000 individuals. Cases and controls for this study will be drawn from this cohort.

To determine the sample size necessary to detect a significant difference in lung cancer risk based on presence or absence of each of the two genetic polymorphisms of interest, we first determined the prevalence of these polymorphisms in the general population by surveying the literature. The GSTM1 null genotype has a frequency of approximately 50% in the general population. (2,8) The GSTTI null genotype has a much lower prevalence rate, about 16% (10,11); we therefore used the GSTTI null genotype prevalence rate in our calculation of sample size to ensure that our study would be adequately powered for the less prevalent risk factor. Furthermore, because the frequency of the outcome of interest, i.e. lung cancer, is very low, and the number of controls available (about 9730) is large, we elected to set the ratio of cases to controls at 1:3 in order to increase power. Because the process of cancer susceptibility is complex and because most previous studies have found odds ratios below 2, we set our minimal detectable OR at 1.7.

To determine the expected proportion cases exposed to the risk factor (i.e., GSTTI null genotype), we used the following formula: (12)

$$\frac{P'}{1-p'} \quad \text{or} \quad \frac{p}{1-p}$$

where p' = proportion of cases exposed to the risk factor
p = proportion of controls exposed to the risk factor
OR = the lowest odds ratio worthy of detection

As above, we know that the frequency of the GSTTI null genotype in the general population is 16%. Therefore, using OR = 1.7 and p = 0. 16 in the equation above, we learn that p' = 0.25.

We can then use the following equation to calculate the sample size required to compare proportions in 2 groups of equal size:

$$\frac{p_1q_1+ p_2q_2}{efFect^2} \quad \frac{2}{effect} \quad + 2 \qquad n=8$$

where p_1, p_2 are the proportions in the 2 groups
q= 1-p
effect = the difference in p

Thus, using p_1 = 0. 16, q_1 = 0.84, p_2 = 0.25, q_2 = 0.75, effect = 0.59, we calculate n to be 342 (i.e., 342 cases and 342 controls).

We then use the following equation to determine the number of cases needed when the ratio of controls to cases is changed from 1: 1 to 3:1. (12)

$$m= \frac{(r + 1)n}{2r}$$

Where m = number of cases
r ratio of controls to cases
n number in each group assuming samples of equal size

Using r = 3 and n = 342 yields m = 228 cases. Thus, we need 228 cases and 228 x 3 = 684 controls. Based on the ELCAP study, in which 2.7% of subjects screened were found to have malignant nodules, we would have to screen 8444 patients to yield 229 cases. Thus the NY-ELCAP cohort of 10,000 patients is clearly adequate for our purposes.

In order to account for important known risk factors, cases and controls will be matched for each of the following factors: age, gender, ethnicity, and pack-years of smoking. Following laboratory determination of genotype for all subjects, odds ratios will be calculated for each of the genotypes, GSTM I null and GSTT I null, as well as the two combined. To determine the role of other factors for which cases and controls were not matched, i.e. family history, age at initiation of smoking, occupational exposures (e.g., asbestos), we will perform multiple logistic regression. As part of a secondary analysis, we may determine odds ratios for various sub-groups of subjects (i.e. based on age, gender, level of smoking) in order to further study the effect of these various risk factors in relation to genotype and lung cancer risk.

C. Study Procedures

Referral and entry. Referrals will be taken by the study coordinator through a dedicated "800 number" phone line. The coordinator will set up an appointment for the screening CT, including evenings

and weekends at the patients' convenience. Baseline demographic data and source of referral will be recorded on structured forms and transmitted directly to the central data repository via a dedicated personal computer. At the appointment, the entire study will be explained to the subject and informed consent obtained. Once consent is obtained, the study coordinator will complete the background questionnaire by interviewing the subject. A phlebotomist will then obtain 20 cc of blood from the subject. Sputum, urine, and oral cell samples will also be obtained. The study coordinator will then guide the subject to the CT scanner.

Baseline CT screening. The CT images will be obtained using a Siemens Somatom Plus4 Spiral CT scanner at low-dose setting. Contiguous I 0-mm-thick slices at endinspiration from thoracic inlet to the adrenal glands will be obtained. For subjects seen at CPMC, images will be interpreted on the Kodak Picture Archiving and Communications System (PACS) by one of three dedicated chest radiologists within 24 hours of the study. Similar systems will be used at the other participating centers. Positive findings will be recorded on structured forms that will be transmitted to the central data repository via a dedicated computer. A written report will automatically be generated. A second reading will be done independently at another one of the participating centers, and findings will be transmitted to the central data repository. Any discrepancies will be discussed between the readers, and consensus readings will be recorded.

Diagnostic work-up on screen-detected nodules. Each subject with 1-6 concalcified nodules will have a standard CT scan of the chest with thin sections through the nodule(s). Standard radiographic criteria will be used to assess the benignity of the lesion. If the lesions cannot be classified as benign, a course of empiric antibiotics will be recommended. If resolution is incomplete, subsequent evaluation will follow standard practice:

1. All nodules < 5 mm in diameter will be followed by limited CT scan after 3 motnhs, and if no growth is observed, at 6, 12, 24, and 36 months. As long as there is no change in size, the nodule will be considered benign.
2. All nodules 5 -10 min in diameter will be evaluated for the possibility of obtaining biopsy by fine needle aspiration, bronchoscopy, and/or video-guided throracoscopy. If biopsy is not possible, the lesion will be followed with CT as described above.
3. All nodules > I cm. in diameter will be biopsied with CT-guided fine needle aspiration, bronchoscopy, and/or video-guided thoracoscopy.

The results of all tests will be documented on structured forms and sent to the central data repository. These results and the above recommendations will be communicated to the referring physicians, although final decisions regarding work-up will be left to the referring physician. Confirmation of diagnosis, by histology or long-term CT follow-up, will be required for each patient.

Treatment. All patients with a noncalcified nodule for which cytology or histology shows malignant cells will be referred for thoracotomy and standard surgical treatment. All pathological specimens will be submitted for pathological review, and results will be documented on structured forms and sent to the central data repository.

Follow-up. Follow-up information on survival of cases will be obtained and documented.

Processing and storage of human samples. Twenty mL of blood will be obtained from each subject by a trained phlebotomist prior to CT scanning. Bar code labels with sample numbers will be used for identification of samples. Sample tracking forms will be completed and shipped with specimens from all centers to the Biorepository Core Facility at the Herbert Irving Cancer Center Core Facility at Columbia.

At the Core Facility, blood samples will undergo low-speed centrifugation to remove plasma, which will then be aliquoted into 1.8 mL vials. The white and red blood cells will be diluted with PBS and then layered over ficoll for centriftigation. Mononuclear cells will be collected and washed with PBS before freezing in edia containing FCS and DMSO to maintain cell viability. A total of 5 aliquots will be made for each sample. Granulocytes will also be collected and washed with a red blood cell lysing buffer and frozen in I mL of Tris-EDTA buffer.

A Microsoft Access computerized database will be used for the sample inventory and will contain information on sample ID, sate of collection, date of prcoessing, processing technician, total volume, types of aliquots, volumes of each aliquot, and location of each vial. As samples are retrieved for analysis, the database will be updated to indicate sample removal, date, and destination. The number of freeze-thaw cycles for each aliquot will also be recorded.

Genotyping for polymorphisms. DNA will be extracted from granulocytes of subjects selected for analysis by a salting out method by Core B at Columbia. Results of all assays will be read from the gels by 2 independent investigators and at least 10% of the samples will be repeated after blinding for quality control. The primers for GSTMI are 5'- GAACTCCCTGAAAAGCTAAAGC-3' and 5'-GTTGGGCTCAAAATCGGTGG-3'.

Primers for beta-globin will be added as an internal control. The presence or absence of the 215-bp band reveals the GSTMI-positive or GSTMl-nuII genotype. Subjects will be classified as gene present (one or two copies) or absent. Approximately 50% of controls will have the null genotype for GSTM I. A similar process will be used for determination of GSTTI genotype, for which we expect approximately 16% of controls will have the null genotype.

Urinary cotinine. As a rapid and inexpensive means to confirm current smoking status, urine samples from all subjects will be analyzed by ELISA. Although this method does not give precise quantitation at low levels (<I 0 ng/ml), it has been used successfully in other studies for a similar purpose. Urinary creatinine will be determined to standardize for urine concentration.

Timeline. Analysis of the samples will begin as soon as sufficient numbers of cases have been identified and matched with appropriate controls.

D. Study Drugs

Not applicable.

E. Medical Devices

Not applicable.

F. Study Questionnaires

During the initial telephone contact with the study coordinator, basic demographic data and the source of referral will be documented on structured forms directly- entered into a dedicated personal computer into the central data repository via the internet, even if the appointment is subsequently cancelled. Once consent is obtained, the study coordinator will complete the study questionnaire by interviewing the subject. The questionnaire will include the following: age, gender, ethnicity, family history of cancer, smoking history (including age at initiation of smoking, number of packs per day smoked, number of years smoking), occupational exposures.

G. Study Subjects

a. Inclusion criteria

Subjects will be asymptomatic men and women older than 60 years of age with a smoking history of 10 pack-years or greater. They will be recruited from outpatient clinics at 8 centers in New York City and 2 others elsewhere in New York State. At CPMC, for example, 2000 patients will be recruited from the outpatient population served by New York Presbyterian Hospital's Associates in Internal Medicine (AIM) clinic. The racial composition of this population is approximately 54% Latino, 32% African-American, and 14% Caucasian, non-Hispanic.

b. Exclusion criteria

Individuals with a history of cancer (except non-melanotic skin cancer) or who are unable to undergo thoracic surgery will be excluded from the study. Surgical eligibility will be determined by standard clinical assessment, including pulmonary and cardiac function tests, if indicated.

H. Recruitment of Subjects

Subjects will be recruited at all the centers through efforts targeted at their primary care physicians. At CPMC, for example, study subjects will be recruited primarily from the AIM clinic, a cornerstone of clinical care in the Washington Heights and Inwood communities (Northern Manhattan) serving 60,000 patients a year. We will mail letters describing the study to all primary care and pulmonary physicians affiliated with AIM. We will also attempt to increase awareness of the study through dissemination of brochures in the AIM clinic and presentations at medical and surgical grand rounds.

I. Confidentiality of Study Data

At the time of entry, each subject will receive a unique identifier. All demographic data transmitted to the central data repository and stored biological samples will be encoded with the unique identifier. Clinical data as well as contact information for the subject and his/her physician will be kept by the study coordinator in a locked file cabinet.

J. Potential Conflict of Interest

None.

K. Location of the Study

The study will be conducted primarily at CPMC, with subjects drawn from all 10 NY-ELCAP centers, including Columbia's AIM Clinic. Data will be collected from all centers in a central data repository. All biological samples will be stored at Columbia's Core Facility.

L. Potential Risks

Participation in the study has several potential risks. One is radiation, as the CT scan delivers a dose of radiation of approximately 0.3-0.6 rads. Furthermore, because CT scanning is very sensitive for lung pathology, many of the abnormalities detected will be benign. Thus, patients with abnormalities on screening CT scan may require further work-up, such as repeat CT scanning, biopsy, bronchoscopy, and/or video-guided thoracoscopy, to determine whether lesions are benign or malignant. Each of these procedures, although potentially life-saving if they identify an early lung cancer, is associated with risks. Although our work-up algorithm is designed to minimize risk, and the ELCAP study did indeed find that 27 of 28 biopsy specimens were malignant, the work-up itself may represent a significant cost to the patient. Although we expect almost 90% of the nodules identified to be benign, the process of determining benignity itself may cost the patient in missed work days, time, and psychological distress. For a confirmed malignant nodule, chest surgery is likely to be recommended.

Blood collection poses little risk to study subjects, other than the slight pain of blood drawing and the potential for bruising at the site of needle puncture. CT screening appointments may be scheduled on evenings and weekends at the patients' convenience. CT screening will be provided free of charge.

M. Potential Benefits

Patients may or may not benefit directly from the study. Benefits may include early detection of lung cancer, such that the patient's chances of undergoing treatment for cure may be significantly increased. The screening CT scan may also detect other diseases, such as emphysema, and thereby lead to treatment. It is also conceivable that increased awareness might lead patients and their physicians to be more motivated in efforts to achieve smoking cessation. There are potential benefits to society, as identification of biomarkers which would help in the targeting of CT scanning to high-risk individuals would be a significant contribution.

N. Alternative Therapies

Patients not participating in the study could elect not to undergo CT scanning of the chest, but instead a standard chest x-ray examination. The radiation dose of a chest x-ray is about 10% of that of a chest CT scan. A CT scan is approximately four times more likely than the standard chest x-ray to detect an early lung cancer.

O. Compensation to Subjects

No monetary compensation will be provided to patients. However, CT screening will be provided free of charge.

P. Costs to Subjects

Subjects will not incur any costs as a result of participating in the study. However, should a cancerous nodule be identified, standard treatment will be offered, and patients will be responsible for the costs of such treatment, either directly or through their medical insurance and/or other forms of medical coverage.

Q. Minors as Research Subjects

The study will include only adult patients of 60 years old or greater.

R. Radiation or Radioactive Substances

Low-dose CT screening delivers a dose of radiation of approximately 0.3-0.6 rads, which is equivalent to about 10 times the dose of a standard chest x-ray examination or one year of background radiation in the greater New York City area. Although this level of radiation is within federal guidelines, radiation risk is cumulative over a lifetime.

S. References

1. Henschke Cl, McCauley DI, Yankelevitz DF, Naidich DP, McGuiness G, Miettinen OS, Libby DM, Pasmantier MW, Koizumi J, Altorki NK, Smith JP. Early Lung Cancer Action Project: overall design and findings from baseline screening. *Lancet* 1999; 354:99-105.
2. Spivack SD, Fasco MJ, Walker VE, Kaminsky LS. The molecular epidemiology of lung cancer. *Crit Rev Toxicology* 1997; 27(4):319-65.
3. Hecht S S. Tobacco smoke carcinogens and lung cancer. J *of National Cancer Institute* 1999; 91(14):1194-1210.
4. Bennett WP, Alavanja MC, Blomeke B, Vahakangas KH, Castren K, Welsh JA, Bowman ED, Khan MA, Flieder DB, Harris CC. Environmental tobacco smoke, genetic susceptibility, and risk of lung cancer in never-smoking women. J *National Cancer Inst* 1999; 91:2009-14.

5. Kato S, Bowman ED, Harrington AM, Blomeke B, Shields PG. Human lung carcinogen-DNA adduct levels mediated by genetic polymorphisms in vivo. J *National Cancer Inst* 1995; 87:902-7.
6. Wang Y, lchiba M, lyadomi M, Zhang J, Tomokuni K. Effects of genetic polymorphism of metabolic enzymes, nutrition, and lifestyle factors on DNA adduct formation in lymphocytes. *Industrial Health* 1998; 36:337-46.
7. Grinberg-Funes RA, Singh VN, Perera FP, Bell DA, Young TL, Dickey C, Wang LW, Santella RM. PAH-DNA adducts in smokers and their relationship to micronutrient intake and serum levels and glutathione-S-transferase MI genotype. *Carcinogenesis* 1994; 15:2449-54.
8. Stucker I, de Waziers 1, Cenee S, Bignon J, Depierre A, Milleron B, Beaune P, Hemon D. GSTM 1, smoking and lung cancer: a case-control study. *Int* J *Epid* 1999; ~28(5):829-35.
9. Houlston RS. Glutathione S-Transferase M I status and lung cancer risk: a metaanalysis. *Cancer Epid, Biomarkers & Prevention* 1999; 8:675-82.
10. Jourenkova N, Reinikanen M, Bouchardy C, Husgafvel-Pursiainen K, Dayer P, Benhamou S, Hirvonen A. Effects of glutathione-S-transferases GSTMI and GSTTI genotypes on lung cancer risk in smokers. *Pharmacogenetics* 1997; 7:515-8.
11. Saarikoski ST,Voho A, Reinikainen M, Anttila S, Karjalainen A, Malaveille C, Vainio H, Husgafvel-Pursiainen K, Hirvonen A. Combined effect of polymorphic GST genes on individual susceptibility to lung cancer. *Int* J *Cancer* 1998; 77:516-2 1.
12. Armitage P, Berry G. Statistical Methods in Medical Research. Blackwell Scientific Publications: Oxford, 1994; 202-204.