Smoking effect on DNA methlyation in peripheral mononuclear cells

ICCR-IRB project
Keith Brenner PGY2
Principal Investigator: Charles Powell, MD

A. Study Purpose and Rational

Lung cancer is the leading cause of cancer death in the United States, accounting for approximately 160,000 deaths annually. Most lung cancer is detected at an advanced stage resulting in a poor 5-year survival of 15% which has not significantly improved in recent years.[1] Since 87% of lung cancer cases are attributable to smoking, it is intuitive to screen for lung cancer in these high risk individuals.[2] However, there is no current recommended screening program for current or former smokers, though multiple noninvasive screening modalities are being investigated. This study examines epigenetic alterations of mononuclear cell DNA in the peripheral blood and the association with smoking exposure. Specific epigenetic profiles in these cells may parallel carcinogenic changes in the lung, thus ultimately serving as a noninvasive screening tool for lung cancer and other diseases associated with epigenetic alterations.

Epigenetic processes, including DNA methylation, are essential for normal development and differentiation, while perturbations in this process via aging and environmental exposures have been implicated in disease.[3] Patterns of DNA methlyation are altered in neoplasia; hypermethylation of promoter regions causes inappropriate gene silencing while regional hypomethylation may lead to genomic instability. [4] Hypermethlyation of CpG islands within promoter regions of tumor suppressor genes is a known epigenetic change in the pathogenesis of lung cancer.[5] Sputum studies have demonstrated methylation of tumor suppression genes are strongly associated with lung cancer risk and smokers are more likely to harbor these hypermethylated regions than nonsmokers.[6] Likewise, an analysis of bronchoalveolar lavage specimens demonstrate those with lung cancer carry a higher frequency of methylation of these genes when compared to smoking controls.[7] These studies suggest that smoking results in promoter hypermethylation, contributing to increased susceptibility of carcinogenesis throughout the lung.

While the above studies demonstrate the ability to identify those at risk for lung cancer based on methylation analysis, sputum expectoration is often inadequate and bronchoalveolar lavage remains an invasive and expensive technique. An alternative is to study methylation patterns in peripheral mononuclear cells, a specimen easily obtained by phlebotomy. Thiagalingam, et. al showed that peripheral lymphocytes contained some of the same smoking induced methylation changes as bronchial epithelial specimens.[8] This suggests that peripheral mononuclear cells may be used as a surrogate to detect alterations of pulmonary DNA methylation patterns in smokers.

In the first phase of this retrospective cohort study, we will determine if smoking exposure is associated with DNA methylation content in peripheral blood mononuclear

cells using two distinct high throughput screens for methylation. The confirmed and validated results will then be used in future studies to determine an association with lung cancer.

B. Study Design and Statistical Analysis

The purpose of this retrospective cohort study is to determine if DNA methlyation in peripheral mononuclear cells is associated with smoking exposure. We will use existing biological specimens obtained during recruitment of 557 subjects into the Early Lung Cancer Action Project (ELCAP).[13]These subjects were current or former smokers without known lung cancer, who were recruited to undergo lung cancer screening with low dose chest cat scans. Enrollment also required the completion of questionnaires, pulmonary function tests, and collection of biospecimens. Peripheral blood of the 557 subjects was used to isolate mononuclear cell genomic DNA, which was the source of investigation in this study.

The outcome of this study is the presence of methylation at 58,000 genetic loci, a survey which evenly covers the genome, including CpG islands. Global analysis of DNA methylation will be performed using two complimentary and unbiased methods. Methylation-sensitive single nucleotide polymorphism chip analysis (MSNP) uses Affymetrix SNP arrays to profile genome wide methylation.[9] 3rd generation 250K SNP chips evenly cover the genome, 20% of these SNPs have a local sequence that is targeted by a methylation sensitive restriction enzyme prior to PCR. By comparing the enzyme treated DNA to control, the degree of methylation at each site can be calculated. The second method, CpG Island Microarray, uses a similar strategy of methylation sensitive restriction enzymes to digest genomic DNA. Digested and mock-digested products are ligated, amplified, and labeled with either Cy5 or Cy3. They then undergo hybridization to a Cpg island array containing previously annotated CpG tags. Hypermethlyation is present if signal is preserved in samples having undergone methlyation sensitive digestion.[10]All targets obtained from the MSNP and CpG island array method will be confirmed using Methylight, a methylation sensitive real time PCR technique.[11]

This study will compare two subsets of subjects enrolled in the ELCAP study. Group 1 will contain current smokers with the highest pack-year history of cigarette smoking, while Group 2 will consist of former smokers with the lowest pack-year history of cigarette smoking. The purpose of this grouping is to maximize the differences in potential smoking-related methylation alterations. The initial analysis will not distinguish whether alterations of methylation are related to current smoking status or cumulative smoking exposure. These two groups may differ in age, which would require adjustment given age is a likely confounder.

Statistical analysis will employ the Fisher's exact test to determine if each of the 58,000 loci significantly differs between group 1 and 2. This method is preferable to chi square analysis. The major problem with analyzing such a large number of markers is that many false positives will be encountered and sorting through to find true positives may prove difficult. Bonferonni offers the most conservative method of adjustment when using

multiple comparisons. Using this method, the alpha value would be lowered from 0.05 for each test to (0.05/58000), or 8.6 X 10^{-7}. In microarray analysis and in other realms, this is considered overly conservative and precludes the discovery of true positives and reduces power. An intermediate approach is to accept a previously specified false discovery rate (the q-value). In our studies, a false discovery rate of 5% will be used and the type one error rate was recalculated to 0.001, using the method described by Benjamini and Hochberg.[12]

The power analysis was constructed with the interest of detecting a 2.5-fold difference of methylation proportions between current heavy smokers compared to former light smokers. Based on prior data, it is estimated that the proportion of methylation at most loci in former light smokers is between 0.20 and 0.40. With a sample size of 80 subjects per group, estimated methlyation proportions, and type I error rate of 0.001, the following is an example of several power calculations.

Proportion methylation in former light smokers	Difference	Effect Size	Power
0.20	2.5X	0.3	77%
0.25	2.5X	0.375	95%
0.30	2.5X	0.45	>98%
0.40	2.0X	0.40	98%

A second analysis will attempt to identify regions of the genome that are differentially methlyated. The 58,000 markers will be divided into 1 megabase groups (bins). Each bin will contain approximately nine markers, resulting in 6000 bins for analysis. The Fisher's exact test will again be used and the multiple comparisons will again be adjusted by controlling the false discovery rate to be less than 5%.

The initial data extraction will elucidate differences in methylation patterns between current smokers with high pack year exposure vs former smokers with low pack year exposure among 160 subjects. These differences in methlyation will be due to current smoking status, cumulative smoking exposure, and type one error. In the next step, we will apply these markers on a larger data set of 397 patients. In this data set, methylation status will be generated using the MethyLight assay, looking only at the differentially methylated markers identified in phase one. Smoking exposure will analyzed according to two measures, (1) former vs current smoker (2) pack-years on continuous scale. This set is not independent since subjects at the extreme of smoking exposure were taken from the former and current smoking groups in the first part of experiment. These subjects will not be reanalyzed in this data set; thus this data set will be deficient in former smokers with low pack year exposures and smokers with high pack year histories. However, given the evidence that smoking, if anything, should increase methylation, this selection bias will decrease our ability to detect differences and only the strongest associations will be identified.

The first analysis, former vs current smoker, will use the Fisher's exact test. Adjusting for multiple comparisons will no longer be required given that we are testing predefined hypotheses. This analysis will compare the remaining 143 current smokers to the 254

former smokers. Assuming the proportion of methylated former smokers is 0.2, a=0.05, we will be capable of detecting a difference of 75% in the proportion of methylation among current smokers, (effect=0.15) at 90% power.

Second, a t-test will be used to determine if those with a specific positive methylation marker have a different mean pack-year history exposure than those without methylation at a specific marker. The 397 remaining patients will be used, assuming normal distribution with a standard deviation of 25 pack years. There will be 90% power to detect a difference of a mean of 7-14 pack-years, depending on the proportion of methylation at given marker.

In future aims of this project (not included), the methylation alterations identified here will be compared to epigenetic changes in lung cancer. Epigenetic changes in peripheral mononuclear cells may mirror carcinogenic alterations in the lung; these changes may be utilized to screen those at high risk for lung cancer.

C. Study Procedure

Patients have been previously enrolled as part of the ELCAP study. Peripheral blood has been stored for all individuals and mononuclear cell genomic DNA has been isolated. These samples will then undergo analysis via MSNP and CpG Island Microarray, as previously discussed. Positive results will be confirmed by methylation sensitive real time PCR. All of these studies will take place at the Columbia Presbyterian campus in the labs of Dr. Powell and collaborating partners.

Patients will not require further contact or procedures.

D. Study Drugs

No drugs will be used in this study.

E. Medical Devices

No medical devices will be used in this study.

F. Study Questionnaires

All subjects were required to complete questionnaires upon enrollment in the study. All 557 patients completed a questionnaire with details on demographics and smoking history, including current versus former smoking, pack years, smoking duration and intensity, year of last cigarette, passive smoke exposure, and type of tobacco exposure. [13]

G. Study Subjects

ELCAP enrolled asymptomatic volunteers over 60 who have a greater than 10 pack year history of smoking without a known history of cancer. They must have been considered fit to undergo thoracic surgery. The following enrollment reflects the subjects enrolled at Columbia.

Number	Total 557	Current Smokers 223	Former smokers 334
Age	66	65	67
Gender, % male	51	51	51
Confirmed lung Ca	13	5	8
Current Smoker %	40	100	0
Total pack years	41	45	38
Average duration	37	44	32
Avg. Yr of last cigarette			1985
Lives with smoker	12	18	9
Parents smoked	80	77	81
Prior exposures			
Cigarette	96	96	96
Pipe	31	21	38
Cigar	35	27	40
Chew	4	3	4

(From Dr. Powell)

H. Recruitment of Subjects

All recruiting has been complete. As described, these were asymptomatic volunteers, many were referred by physicians.

I. Confidentiality of Study Data

All study data has been coded and stored in a secure location.

J. Potential Conflict of Interest

There is no conflict of interest to disclose for any investigator involved in this study.

K. Location of Study

All analysis will be performed at the Columbia Presbyterian Medical Center. There will be no transfer of materials for study at other institutions.

L. Potential Risk

The patients have previously given the blood samples used in this analysis. There are no further risks or discomforts to the subject.

M. Potential Benefit

This study offers no potential benefit to the patient. Elucidating the link between epigenetic changes in peripheral mononuclear cells and smoking/lung cancer will be of considerable societal benefit.

N. Alternative Therapies

This study does not involve experimental therapies.

O. Compensation to subjects

There was no compensation offered for enrollment into ELCAP or this study.

P. Costs to Subjects

The subject will not incur any cost as result of this study.

Q. Minors as Research Subjects

All subjects in this study are older than 60 years.

R. Radiation or Radioactive Substances

Subjects will not be exposed to additional radiation or radioactive materials.

S. Consent

All patients have previously consented upon enrollment into the ELCAP study. During this period, the subjects provided samples of blood and previously consented to the use of blood samples for research purposes, including genetic analysis.

Bibliography

1. Jemal, A, Siegel, R, Ward, E, et al. Cancer statistics, 2006. CA Cancer J Clin 2006; 56:106.

2. Wingo, PA, Ries, LAG, Giovino, GA, et al. Annual report to the nation on the status of cancer, 1973-1996, with a special section on lung cancer and tobacco smoking. J Natl Cancer Inst 1999; 91:675.

3. Jaenisch, R. Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signal. Nature Genetics 2003; 33:245.

4. Jones, PA. Baylin SB. The fundamental role of epigenetic events in cancer. Nature Reviews Genetics. 2003; 6:415.

5. Belinsky, SA. Nikula, KJ. Palmisano WA, et al: Aberrant methylation of p16/INK4A is an early event in lung cancer and a potential marker for early diagnosis. Proc Natl Acad Sci. 1998;61:11891.

6. Belinsky SA, Klinge DM, Dekker JD, et al. Gene promoter methlyation in plasma and sputum increases with lung cancer risk. Clin Cancer Res. 2005; 11:6505.

7. Kim H, Kwon YM, Kim JS, et al. Tumor-specific methlyation in bronchial lavage for the early detection of non-small-cell lung caner. J Clin Oncol. 2004; 22:2363.

8. Russo AL, Thiagalingam A, Pan H. Differential DNA hypermethylation of critical genes mediates the stage-specific tobacco smoke-induced neoplastic progression of lung cancer.Clin Cancer Res. 2005; 11:2466.

9. Yuan E, Haghighi F, White S. A single nucleotide polymorphism chip-based method for combined genetic and epigenetic profiling: validation in decitabine therapy and tumor/normal comparisons. Cancer Res. 2006; 66:3443.

10. Heisler LE, Torti D, Boutros PC, et al. CpG Island microarray probe sequences derived from a physical library are representative of CpG Islands annotated on the human genome. Nucleic Acids res. 2005; 33:2952.

11. Eads CA, Danenberg KD, Kawakami K, et al. MethylLight: a high-throughput assay to measure DNA methylation. Nucleic Acids Res. 2005; 28:E32.

12. Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. 1990. Stat Med; 9:811.

13. Henschke CI, et al. Early Lung Cancer Action Project: overall design and findings from baseline screening. Lancet. 1999; 354:9173.