University of Pennsylvania

Annual Progress Report: 2006 Nonformula Grant

Reporting Period

June 1, 2009 – May 31, 2010

Nonformula Grant Overview

The University of Pennsylvania received $4,206,096 in nonformula funds for the grant award period June 1, 2007 through May 31, 2011. Accomplishments for the reporting period are described below.

Research Project: Project Title and Purpose

Center for the Study of Gene-Environment Interactions in Lung Cancer - Lung cancer is the leading cause of cancer deaths in Pennsylvania. Chemicals in tobacco smoke cause lung cancer, but more than a tenth of cases occur in non-smokers. Industrialized urban areas have higher lung cancer rates than areas with good air quality indicating that environmental pollution is also involved. Although genetic factors increase lung cancer risk, most of these remain unidentified. This project will study genetic differences between smokers and non-smokers, with and without lung cancer, who attend the hospitals of the University of Pennsylvania (and other recruitment sites) in Philadelphia (high air pollution) and Penn State University (and other recruitment sites) in Hershey (low air pollution). Variants of genes that interact with cancer-causing chemicals in tobacco smoke and polluted air, and biomarkers of chemical exposure in urine and blood, will be tested for their association with increased risk of lung cancer and for their association with biomarker levels *per se*.

Anticipated Duration of Project

6/1/2007 - 5/31/2011

Project Overview

This program will define gene-environment interactions that increase risk of lung cancer in African American and Caucasian smokers and, if possible, non-smokers. It will also delineate the relationships between key genetic variants and levels of presumptive risk-conferring biomarkers in the subject sub-groups. The initial target was to recruit six hundred subjects (primarily smokers) with lung cancer and the same number of disease-free control subjects through each of two centers: the University of Pennsylvania Health System (UPHS) in Philadelphia serving a population living in an urban, industrial environment with poor air quality; and the Penn State Milton S Hershey Medical Center (PSHMC) serving a rural population living in a less polluted region. At USPHS the initial target was that equal numbers of African Americans and

Caucasians would be enrolled. Philadelphia County has almost 1.5 times the incidence of lung cancer as Dauphin County, in which PSHMC is located.

A candidate gene approach will be taken. The major carcinogens in tobacco smoke are nitrosaminoketones (NNK) and polycyclic aromatic hydrocarbons (PAH). Polluted air also contains high levels of PAH, together with carcinogenic particulates derived from fossil fuel combustion. Variants of genes involved in the metabolic activation and subsequent detoxification of NNK and PAH, as well as DNA adduct repair, are excellent genetic candidates for lung cancer susceptibility. As methylation is important for maintaining a normal cell phenotype, and for the efficient function of the above activating and detoxifying enzymes, variants of genes involved in folate metabolism may act as risk modifiers. Additional genes of interest are those involved in promoting and controlling inflammatory and immune processes, facilitating DNA repair, and determining addiction to nicotine. DNA from all subjects will be subjected to high throughput single nucleotide polymorphism (SNP) genotyping of the above classes of genes using the Illumina Infinium platform (not the originally proposed lower throughput Golden Gate platform). Biomarkers of NNK, PAH and folate will be tested in the urine and blood of the same subjects.

Sophisticated biostatistical tests that account for confounders such as population stratification will be applied to the genetic, biomarker, clinical and environmental exposure data to assess the relationship between key genes and/or biomarkers and lung cancer susceptibility in African Americans, Caucasians, smokers living in relatively polluted locations, smokers living in relatively unpolluted locations, and non-smokers who have the disease. An important parallel and complementary effort will be directed at assessing the genetic underpinning of differential biomarker responses to environmental factors.

Initial Hypothesis and Aims:

Our initial hypothesis was that there are genetically mandated differences in susceptibility to lung carcinogens that are [a] tobacco-specific [e.g. nicotine derived nitrosamino ketone; NNK: 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone], and [b] found in tobacco smoke and the environment (e.g. polycyclic aromatic hydrocarbons, PAH), and that such individual differences result in predisposition to lung cancer. This predicts that (i) lung cancer susceptibility is governed by complex gene-environment interactions with SNPs in particular genes contributing to the "mutator" phenotype that mutates proto-oncogenes or tumor suppressor genes to transform lung cells; and (ii) polluted environments in which PAH (and other carcinogens) are at relatively high levels will act via a subset of susceptibility genes to raise lung cancer incidence in smokers and non-smokers alike. Carcinogens are metabolically activated to electrophiles (ultimate carcinogens), which if not detoxified, form DNA adducts; such adducts, if not repaired, cause somatic mutations. Many of the genes involved in NNK activation (*CYP2A6, CYP3A4*), NNK elimination (*CR1, AKR1C1/AKR1C2, UDPGT1A9, UGT2B7*), and repair of resultant O^6-methylgaunine and O^6-pyridyloxylbutyl-gaunine lesions (*AGT*) are known. Similarly, many of the genes involved in PAH activation (*AhR, AhRR, CYP1A1, CYP1B1, EPHX1, AKR1A1, AKR1C1-AKR1C3*) and detoxication (*GSTM1, GSTP1, GSTT1, NAT, COMT, NQO1, SULT1A1, UGT1A10*), repair of abasic sites (*APE1*), repair of bulky PAH-DNA adducts (*XPC, XPG, XPF*) and repair of PAH-mediated oxidative DNA damage (*hOGG1, XRCC1*) are also known. Sites of DNA-adduct formation, and COMT activity, are impacted by methylation status. Variants in

folate pathway genes (*MTHFR, MTR, MTRR, TYMS, MTHFD1, DHFR, SHMT*) and diet could therefore be involved directly, or as risk modifiers. This proposal aimed to identify candidate gene SNPs that are over-represented in smokers with lung cancer relative to disease free smokers. The findings may be translatable to non-smokers who are PAH-exposed. The pool of candidate genes will be enriched with novel targets selected from current literature and other sources. This knowledge may identify those most at risk of lung cancer in two regions of Pennsylvania with very different levels of air pollution and with different population demographics. The following specific aims were proposed:

Aim#1. Candidate Gene Approach:
DNA from Caucasian and African American smokers (and non-smokers) with lung cancer and disease free controls will be screened for candidate gene SNPs using Illumina Infinium 10,000-plex genotyping assays. Subjects exposed to high and low levels of air pollutants will be recruited through the Penn and PSMC sites, respectively. An analytical component will determine exposure phenotype (e.g. urinary cotinine and 3-hydroxybenzo[*a*]pyrene glucuronide levels will serve as biomarkers for nicotine and benzo[a]pyrene exposure, respectively; urinary NNAL-glucuronide [(4-methylnitrosamino)-1-3-pyridyl-1-butanol (NNAL-GLu)] and *anti*-benzo[*a*]pyrene-diol epoxide (*anti*-B[*a*]PDE)-N^2-dGuo adducts will serve as biomarkers for exposure to the ultimate carcinogens, NNK and *anti*-B[*a*]PDE respectively) using stable-isotope dilution LC-MRM/MS methodologies. Folate phenotypes will be established using recently developed stable-isotope dilution LC-MRM/MS methodologies. Appropriate case-control, within-case, and inter-group analyses will be performed.

Aim#2. Region and race specific environmental and nutritional modifiers in cases and controls:
A questionnaire-based approach will be taken to capture data on exposure to pollutants and potentially protective nutritional variables in cases and controls in order to address disparities in these factors between regions and between demographically distinct groups of cases and controls. Attitudes to the utility of information on exposure to pollutants and estimates of genetic risk will be assessed. Questionnaires will be used to define disparities in prior history of access to medical care, medical facilities and information, particularly the extent to which differences in such factors might underpin clinical status, including stage of disease, at diagnosis.

Aim#3. Internship Training:
Two Lincoln University students who wish to pursue careers in a health care related profession will be selected in each of the four years of the award by competitive application for summer internship training at Penn and PSMC. This program is designed to help prepare students from this historically black college to make informed choices for graduate school and to make successful transition to graduate studies.

Principal Investigator

Alexander S Whitehead, DPhil
Professor of Pharmacology
University of Pennsylvania
153 Johnson Pavilion
Department of Pharmacology

3620 Hamilton Walk
Philadelphia, PA 19104-6084
(215) 898-2332

Other Participating Researchers

Trevor M Penning, PhD, Ian Blair, PhD, Clementina Mesaros, PhD, Steven Albelda, MD, Anil Vachani, MD, Donald Baldwin, PhD, Caryn Lerman, PhD, Wensheng Guo, PhD (replaces Andrea Troxel, ScD), Mingyao Li, PhD, Timothy Rebbeck, PhD – employed by University of Pennsylvania.
Judith Thomas, PhD, Patricia Joseph, PhD – employed by Lincoln University
Philip Lazarus, PhD, Joshua Muscat PhD, Henry Wagner, MD, Kwang Mi Ahn, PhD – employed by Pennsylvania State University.
Margie Clapper, PhD – employed by Fox-Chase Cancer Center
John Travaline, MD – employed by Temple University Medical School (new recruitment site).

Expected Research Outcomes and Benefits

This project will identify genes that interact with the environment, in particular chemicals from tobacco smoke and fossil fuel combustion, to increase the risk of lung cancer in susceptible individuals. It will test whether variants of other genes involved in important nutritionally related processes, nicotine addiction, and inflammatory/immune functions are independently associated with lung cancer risk or modify (increase or decrease) the risk of lung cancer that is conferred by particular gene-environment interactions. In addition, it will delineate the relationships between genetic factors and key biomarkers that are indicative of exposure to environmental carcinogens and inter-individual differences in the elimination or reduction of the most harmful metabolites of such carcinogens.

The identification of genetic factors that increase the risk of lung cancer (either directly or via biomarkers that are indicative of enhanced susceptibility to high carcinogen load) would enhance the quality of the public health information that is disseminated to the general population during cancer awareness and smoking cessation programs. At the level of the individual, genetic tests, perhaps in combination with biomarker measurement, might be developed that would permit screening for those in whom the risk of lung cancer is significantly elevated. Particular effort and resources could be focused on such individuals to encourage them to stop smoking or to participate in lung cancer early detection programs. It might encourage local, state and federal support for the establishment of the latter.

The knowledge that some individuals, smokers or non-smokers alike, are genetically susceptible to lung cancer when exposed to high levels of environmental pollution should galvanize community action groups, particularly those from disadvantaged and minority neighborhoods (in which the highest concentrations of polluting industries, facilities and heavy commercial traffic are often found), to resist pollution-generating development and to seek remediation of existing sources of pollution.

This project might also assist in the future development of genetic or biomarker-based tests for

those with lung cancer that would allow a treating physician to understand the underlying processes that had led to disease, to predict the likely course of cancer progression, and even to choose the most appropriate therapy.

Summary of Research Completed

Subject Recruitment through June 30, 2010.

Penn: Full IRB approval has been maintained to date.

Cumulative accrual at Penn comprised 472 patients with confirmed lung cancer and 345 controls (Table 1); 176 and 107 of each respective class have been recruited during the current reporting period.

Fox Chase Cancer Center (FCCC): Full IRB approval has been maintained to date.

In 2009 FCCC was added to the program as a recruitment site under the direction of Margie Clapper, Ph.D. Cumulative recruitment comprised 111 patients with confirmed lung cancer and 60 controls (Table 2).

Temple University: Temple University Medical School has been added as a recruitment site to bolster numbers of African American cases and controls. IRB approval is in place and recruitment is under way. However, numbers of enrollees are not yet available.

Enhanced recruitment of African American cases and controls: African American recruitment has been below target and recruitment and sample acquisition efforts have consequently been expanded. We have identified the following sources of study subjects and have projected the yield going forward (i.e. in addition to those already enrolled): VA hospital (45); Hospital of the University of Pennsylvania (25); Fox Chase Cancer Center (30); Temple University (55); recently archived samples from another study conducted by Dr. Vachani (50); and archived samples from another study conducted by Dr. Lazarus at Temple University (50). A conservative projected total yield is 315.

PSHMC: Full IRB approval has been maintained to date.

Cumulative accrual at PSHMC comprised 228 patients with confirmed lung cancer and 306 controls (Table 3); 99 and 131 of each respective class have been recruited during the current reporting period.

The above recruitment total includes subjects recruited through Pinnacle Health Systems. Recruitment at the Hershey sites remains below target. However, expansion of recruitment criteria, once approved by the local IRB, should lead to an increase in recruitment going forward. It is anticipated that the number of cases and controls that will be enrolled by the end of the recruitment phase will be sufficient to permit the main analytic aspects of the project to be undertaken.

Sample and Data Collection.

Sample collection is proceeding at all active sites with batch transfers of biological specimens to the central processing/biomarker facility at Penn.

Data collection is proceeding at all active sites with effective data storage and transfer to the central database.

Measurement of Folates and Homocysteine.

We have capitalized on our development of robust, reproducible liquid chromatography-multiple reaction monitoring/mass spectrometry (LC-MRM/MS) methods for measuring key folate derivatives (5-methyl-tetrahydrofolate, tetrahydrofolate and 5,10-methenyl-tetrahydrofolate) in red blood cells and plasma, and homocysteine (Hcy) in plasma to establish that folate and Hcy concentrations are stable in healthy individuals over time (Mitchell et al, Clin Biochem 2009) but are modified in response to environmental stressors (i.e. drugs) that affect folate-related enzymes (unpublished). Furthermore we have published a second paper supporting relationships between folate-related genotypes and phenotypes and circulating concentrations of the potent pro-inflammatory chemokine MCP-1 (Hammons et al, Clin Immunol 2009). The latter finding reinforces the likely relationships between folate/Hcy metabolism and inflammation and mandates further study within the program.

Our novel LC-MRM/MS methodology has made it possible to identify four distinct (arbitrarily defined RBC folate phenotypes in subjects with defined *MTHFR*) 677C>T genotypes.

The Type I and Type IV phenotypes represent biological extremes (*per se* and with respect to different cellular potential to support the various folate dependent downstream processes). They are predominantly, though not exclusively, associated with the MTHFR 677CC and TT genotypes respectively; hence, there are 12 genotype/phenotype combinations with some being counterintuitive. These will be of considerable interest in this program and will be tested as composite predictors of lung cancer *per se*, age of onset etc. Subdivision of genotype-phenotype combinations will allow the analysis of phenotype as a categorical variable, with different cut-points (i.e. subdivision of phenotypes into "Types" according to readout criteria coincident with, or distinct from, those in figure 1), or with different combinations of genotype-phenotype classes according to their theoretical biological potential and relatedness.

To date, we have characterized composite serum folate and red blood cell folate phenotypes (using the above methods) in 189 cases and 7 controls. DNA from these individuals has been subjected to conventional genotyping for key polymorphisms in MTHFR, CBS, MTR and MTRR. Once data from 200 cases and 200 controls have been accrued, biostatistical analyses will be undertaken. This will serve as a pilot study to "field test" procedures whereby laboratory (genotype and phenotype) data is brought together with *in silico* data to facilitate the investigation of gene-environment interactions in the context of clinical status. This phase of the project has particular relevance given recently published work suggesting that folic acid supplement use is associated with an increase in lung cancer incidence (Ebbing M et al, JAMA 2009).

Transcriptomic Consequences of Changes in Folate Phenotype.

Given the recent interest in the literature regarding the role of folate status in oncogenesis we have subjected cultured cells to growth media containing high and low concentrations of folic acid and to various exposures to anti-folate drugs in order to elicit transcriptomic changes that may validate and/or extend our list of candidate genetic factors. The data are currently being evaluated; however, a disproportionate number of immune related genes appear to be differentially regulated by "folate stress" thereby reinforcing the rationale for their inclusion in our genotyping platform design.

Biomarker Methods Development and Implementation.

Technical details regarding assay development were included in the last report.

The following classes of biomarkers are under consideration for inclusion in the program. The measurement error in these assays is very small thereby permitting genotype/biomarker studies to be undertaken using only a subset of samples: this should facilitate analysis with sufficient statistical power to be undertaken in selected comparative groups (i.e. cases/controls, Caucasians/African Americans, and possible smokers/non-smokers etc.)

DNA-adducts

- Urinary 8-oxo-dGuo (oxidative DNA damage)
- Urinary heptanone-etheno-dCyd (lipid peroxide-mediated DNA damage)
- Urinary B[*a*]PDE-dGuo-adduct (in development)

Metabolites

- RBC 5-methyltetrahydrofolate
- RBC 5,10-methylenetetrahydrofolate
- RBC-tetrahydrofolate

Lipids

- Urinary HETEs
- Urinary 8-iso-$PGF_{2\alpha}$
- Urinary $PGF2_{\alpha}$

Nicotine and metabolites

- Urinary nicotine and nicotine glucuronide
- Urinary cotinine and *trans*-3'-hydroxy-cotinine and glucuronide

NNK and metabolites

- Urinary free NNAL
- Urinary O- and N-NNAL glucuronide metabolites

PAH metabolites

- Urinary 3-hydroxy-B[*a*]P (in development)
- Urinary 1-hydroxy-pyrene

Development of a List of Candidate Genetic Risk Factors for Lung Cancer and Manufacture and Implementation of Platform

Due to changes in cost structure for Illumina Genotyping Products, the genotyping platform will be Illumina Infinium rather than the originally proposed Illumina Golden Gate.

This permits an increase in SNP content of more than 6-fold to around 10,000 data points.

The SNP content design has been finalized and will incorporate approximately 250 genes in the pathways involving:

- Nitrosaminoketone (NNK) activation and detoxification
- Polycyclic Aromatic Hydrocarbon (PAH) activation and detoxification
- Repair of NNK-mediated and PAH-mediated DNA damage
- Nicotine addiction
- Folate and homocysteine metabolism
- Inflammatory signaling and processes

Manufacture by Illumina is currently under negotiation. The design will form the foundation for an off-the-shelf genotyping platform for future studies of smoking and air pollution associated cancers and other chronic diseases.

Student Internship Program

The final two student interns from Lincoln University were selected in February 2010 and undertook training at Penn and PSMC between June and August 2010.

Broadly the training included the following:

- Certification in Human Subjects Research.
- Shadowing clinicians in practice and during the patient-oriented aspects of this project
- Training in questionnaire and interview techniques.
- Training in data entry and management.
- Practical laboratory training in genotyping and other techniques.
- "Hands-on" administration of questionnaires and interviews.
- Attendance at seminars, training sessions, lab/group meetings and chalk talks.
- Reading assignments aimed at acquiring a working knowledge of the background literature for the research.
- Preparation of report(s) detailing work undertaken.
- Formal oral presentation at Lincoln University.

The interns in all four years were selected from strong applicant pools. Most had declared an interest in attending Medical School and feedback from the students themselves indicates that this interest has been strengthened and focused with respect to possible post-MD clinical career path.

Lincoln University Questionnaire-Based Student Study.

Dr. Patricia Joseph intends to conduct a questionnaire-based study of smoking history, exposure and attitudes among Lincoln University students. To this end, Dr. Joseph has worked with Dr. Amy Lazev of FCCC to devise a method for surveying LU students using as a basis, the "College Student Situational Smoking Survey" developed by Dr. Lazev and her colleagues. This initiative is described more fully in the previous report.

Development of a New Course at Lincoln University.

Dr. Patricia Joseph developed a new Honors Course "Environmental Justice" at Lincoln University. This course was given in the Fall Semester 2009 at 5.00-7.30 pm each Monday. Approximately 12 students enrolled.

Publications related to this project:

Summers CM, Mitchell LE, Stanislawska-Sachadyn A, Baido SF, Blair IA, Von Feldt JM and Whitehead AS. Genetic and lifestyle variables associated with homocysteine concentrations and the distribution of folate derivatives in healthy premenopausal women. *Birth Defects Research (Part A) Clin Mol Teratol*, 2010; 88: 679-688.

Truong T, Hung RJ, Amos CI, Wu X, Bickeböller H, Rosenberger A, Sauter W, Illig T, Wichmann HE, Risch A, Dienemann H, Kaaks R, Yang P, Jiang R, Wiencke JK, Wrensch M, Hansen H, Kelsey KT, Matsuo K, Tajima K, Schwartz AG, Wenzlaff A, Seow A, Ying C, Staratschek-Jox A, Nürnberg P, Stoelben E, Wolf J, Lazarus P, Muscat JE, Gallagher CJ, Zienolddiny S, Haugen A, van der Heijden HF, Kiemeney LA, Isla D, Mayordomo JI, Rafnar T, Stefansson K, Zhang ZF, Chang SC, Kim JH, Hong YC, Duell EJ, Andrew AS, Lejbkowicz F, Rennert G, Müller H, Brenner H, Le Marchand L, Benhamou S, Bouchardy C, Teare MD, Xue X, McLaughlin J, Liu G, McKay JD, Brennan P, Spitz MR. Replication of lung cancer susceptibility loci at chromosomes 15q25, 5p15, and 6p21: a pooled analysis from the International Lung Cancer Consortium. *J Natl Cancer Inst* 2010; 102 :959-71. Epub 2010 Jun 14.

Balliet RM, Chen G**,** Dellinger RW, Lazarus P. UGT1A10: Activity against the tobacco-specific nitrosamine, NNAL, and a potential role for a novel UGT1A10 promoter deletion polymorphism in cancer susceptibility. *Drug Metab Dispos* 38:484-90, 2010.

Table 1. University of Pennsylvania (Cumulative through 06/30/10)		
	Lung Cancer	**Controls**
Gender		
Male	269	177
Female	203	168
Race		
White	384	203
African-American	79	134
Asian	7	6
Other/Unknown	2	2
TOTAL	**472**	**345**

Table 2. Fox Chase Cancer Center (Cumulative through 06/30/10)		
	Lung Cancer	**Controls**
Gender		
Male	59	7
Female	52	53
Race		
White	95	50
African-American	12	3
Asian	0	4
Other/Unknown	4	2
TOTAL	**111**	**60**

Table 3. Penn State Hershey Medical Center (Cumulative through 6/30/2010)		
	Lung Cancer	**Controls**
Gender		
Male	133	146
Female	95	160
Race		
White	212	264
African-American	12	35
Asian	0	0
Other/Unknown	4	7
TOTAL	**228**	**306**