University of Pennsylvania

Annual Progress Report: 2006 Nonformula Grant

Reporting Period

June 1, 2010– May 31, 2011

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*.

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 nicotine derived 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 (*CBR1, 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 is 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 adduct)s 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

Progress through 5/31/11

Subject Recruitment.

Recruitment at all sites was terminated by 02-28-11, to allow sufficient time for genotyping, completion of biomarker assays and data analysis. Cumulative accruals of all consenting lung cancer cases and controls are given in Tables 1-4. The original target recruitment was 600 lung cancer cases and controls from urban Philadelphia and 600 lung cancer cases and controls from rural PA (Hershey). The number of lung cancer cases and controls was adjusted down to 300 lung cancer cases and controls at Hershey in 2008. Final recruitment numbers were: 732 and 501 lung cancer cases and controls from urban Philadelphia (Penn, Temple and FCCC combined) and 278 and 362 lung cancer cases and controls from rural PA (Hershey). Recruitment completed in the last year of the award is summarized below.

University of Pennsylvania (Penn): Full IRB approval is still in effect. Final recruitment at this site comprised 571 patients with confirmed lung cancer and 398 controls (Table 1); 99 and 53 of each respective class were recruited during the current reporting period.

Temple University: Full IRB approval is still in effect. Temple University Medical School was added as a recruitment site in 2010 under the direction of Dr. John Travaline to bolster numbers of African American cases and controls. 13 lung cancer cases and 4 controls were recruited. These totals have been combined with the Penn totals in Table 1 to give 584 lung cancer cases and 402 controls.

Fox Chase Cancer Center (FCCC): Full IRB approval is still in effect. In 2009 FCCC was added to the program as a recruitment site under the direction of Margie Clapper, Ph.D. Final recruitment at this site was 148 patients with confirmed lung cancer and 99 controls (Table 2). 37 and 39 of each respective class were recruited during the current reporting period.

Enhanced recruitment of African American cases and controls: African American recruitment was targeted at 300 lung cancer cases and 300 controls from urban Philadelphia. Final recruitment numbers were 122 lung cancer cases and 144 controls. These numbers were enriched from archived samples from other studies performed by Dr. Vachani (Penn) and archived samples from another study conducted by Dr. Lazarus at Temple University (Table 3). When combined with African Americans from the current study we achieved 243 lung cancer cases and 222 controls providing a total of 465 African Americans for genotyping (Tables 1–3).

Penn State Hershey Medical Center (PSHMC): Full IRB approval has been maintained to date. Final recruitment at PSHMC comprised 278 patients with confirmed lung cancer and 362

controls (Table 4); 50 and 56 of each respective class were recruited during the current reporting period. The above recruitment total includes subjects recruited through Pinnacle Health Systems which was added to bolster recruitment at this site.

Sample and Data Collection.

All biospecimens were transferred to the biomarker core at Penn; all bloods were processed for DNA in the genotyping core and samples quality controlled. Completed clinical data and questionnaire information was batch downloaded into the Oracle Clinical Data Base designed and maintained in the Clinical Computing Research Unit at Penn. For each subject recruited we have a complete clinical record, smoking history and other demographic information (age, gender, and address, etc). Biomarker and genotyping data were maintained in separate data bases due to the large volume of data generated.

Biomarker Methods Development and Implementation.

Technical details regarding assay stable isotope dilution liquid chromatography mass spectrometric assays were detailed earlier. The following biomarkers were measured in lung cancer cases and controls from the Penn and Hershey cohorts in either urine or blood as specified, see Table 5. [1] 8-oxo-dGuo in urine (biomarker of oxidative stress and oxidative damage to the nucleotide pool); [2] Nicotine and its 15 metabolites (biomarker of smoking use and behavior); [3] Total NNAL (biomarker of NNK exposure a tobacco specific carcinogen); [4] Total 1-OH-pyrene (biomarker of exposure to PAH); [5] Creatinine as a denominator for the urinary biomarkers; [6] measurement of blood folates as: 5-MTHF in RBC; THF in RBC; 5,10-MFTHF in RBC; folic acid in RBC; 5-MTHF in plasma; THF in plasma; 5,10-METHF in plasma and folic acid in plasma. In total > 7300 samples were analyzed and a total of 17,744 analytes measured.

Genotyping

(i) *Genotyping Chip Design.* A custom lung cancer genotyping chip was designed, refined and built on the Illumina Infinium platform in the fall/winter of 2010. The final design included 250 candidate genes with 12,649 SNPs including 380 Ancestry Informative Markers. To assist the design process we created a new web-based analysis tool, Tagger Batch Assistant, which has been released for public use. This software allows large sets of target genes to be submitted for automated analysis by the Tagger algorithm for predicting haplotype tag SNPs, and manages the output data in formats convenient for transfer to assay design tools. Using Tagger Batch Assistant, we determined that the Tagger Multimarker algorithm produced genetic coverage of the target genes that nearly equaled that of the more conservative Pairwise algorithm while requiring significantly fewer tag SNPs. Tagger Multimarker SNPs were collected from the CEU (Caucasian) and YRI (Yoruban; surrogate for African American) HapMap population databases, filtered for tagging correlation R^2 of 0.8 or higher and MAF equal to or greater than 1%, and processed through the online Illumina design tools to calculate a design score predictive of assay success. A design score threshold of 0.6 or higher was applied, and the resulting marker SNP set was combined with a list of non-synonymous coding SNPs identified as influencing the activities of the targeted genes. Candidate genes that were in the final panel were from pathways that may

influence the biomarkers measured: (i) genes involved in oxidative stress; (ii) genes involved in nicotine metabolism; (iii) genes involved in nitrosaminoketone (NNK) activation and detoxification; (iv) genes involved in polycyclic aromatic hydrocarbon (PAH) activation and detoxification; (v) genes involved in nucleotide excision repair; (vi) genes involved in base excision repair; (vii) genes involved in translesion by-pass; (viii) genes involved in estrogen metabolism; (ix) genes involved in folate and homocysteine metabolism; and (x) genes involved in inflammatory signaling and processes.

(ii) *Genotyping.* Blood and buffy coat samples were extracted using a Chemagen robot and Chemagen magnetic bead kits for DNA purification. DNA extractions produced an average of 15.8 ug per sample. UV spectrophotometry was used to measure DNA concentration and purity, and DNA integrity was confirmed by gel electrophoresis on every sample. DNA from 1889 primary samples and 44 randomly chosen duplicates was genotyped; negative control wells containing water produced no assay results indicating low likelihood of cross-contamination.

The Illumina Infinium genotyping assay was performed according to the manufacturer's protocol. Infinium internal quality control metrics were normal, and the genotyping calls were assigned using the standard Illumina clustering algorithms.

Analysis of Genotyping Data.

(i) Quality Control. The original genotype dataset contained 1889 unique individuals. We performed quality control measures to exclude unreliable samples. First, samples were excluded if the genotype call rate was < 97.5% or if the sample showed excess or deficient heterozygosity (n = 51) (Figure 1). Next, we assessed for cryptic relatedness or erroneous duplicates among participates using identical by decent estimation (IBD) (n = 96) (Figure 2). Third, we removed water samples and those with mismatched IDs (n = 22). After these eliminations, we then conducted multi-dimensional scaling (MDS) analysis to determine subjects' genetic ancestry (Figure 3). We used this information to identify genetically inferred Caucasians and African Americans. After merging with clinical phenotype data, we have 1094 Caucasians and 433 African Americans in the final analysis. We also performed quality control measures to exclude unreliable SNPs. For the analysis reported here, we eliminated SNPs with genotype call rate < 95%, with minor allele frequency (MAF) < 5% or if there was significant departure from Hardy-Weinberg equilibrium ($P < 0.001$).

(ii) Genetic Association Analysis (All Caucasians). Our primary analysis compared genotype frequencies in lung cancer cases with frequencies in controls adjusting for age, gender, study site and smoking status (Figure 4). All analyses were conducted using PLINK software. After association analysis was conducted, we made a Q-Q plot to compare the observed P-value distribution with that expected under the null distribution that assumes no association, no population stratification, and genotyping bias, and calculated the genomic control inflation factor λ, defined as the median of the observed 1-df chi-squared statistics divided by 0.455 (Figure 5). The λ value for Caucasians is 1.0, indicating that there is no obvious population stratification in our data. Table 6 shows the detailed association results for SNPs with P-values less than 0.001. Four SNPs in the *CYP2E1* gene (rs743535, rs7092584, rs11593189 and rs8192772) showed a highly significant association with Caucasian lung cancer incidence. *CYP2E1* plays a role in the

metabolic activation of the tobacco specific nitrosamines (*N*-nitrosodimethylamine, *N*-nitrosonornicotine, NNK and NNAL), and the metabolism of the environmental pollutant benzene. We also began association of SNPs in pathway specific genes with lung cancer incidence in Caucasians. The genotype counts of MTHFR 677C>T in cases and controls of the Caucasian samples are summarized in Table 7. The association tests on the SNPs within *MTHFR, FPGS, MTR, MTRR, CBS, SHMT1, SHMT2, TYMS, DHFR, TCN2* showed no striking significance with lung cancer incidence.

Student Internship Program

Both Penn and PSHMC each hosted a summer intern from Lincoln University from 06/10 - 08/10. The internship experience was identical to that reported previously.

Honors Environmental Justice 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. The EJ course served as a template for other related Honors courses. Notably during the spring 2011, the course, Healthcare Disparities (HON300) taught by Dr. Helen Major of the Science division, was modeled after the EJ course and was made possible as a result of the experiences gained from working with this grant. Twenty-one students were enrolled and successfully completed the course.

Publications related to this project:

Prior to current reporting period.

Chen G, Dellinger R, Gallagher CJ, Sun D, Lazarus P. Identification of a novel polymorphism in the UGT2B10 gene. Loss of function of the codon 67 (Tyr) variant in tobacco-specific nitrosamine glucuronidation. *Pharmacogenet Genomics* 2008, 18:189-191.

Chen G, Dellinger R, Sun D, Spratt TE, Lazarus P. Glucuronidation of 4-(methyl-nitrosamino)-1- (3 pyridyl)-1-butanol (NNAL) and other tobacco-specific nitrosamines by UGT2B10. *Drug Metab Dispos*. 2008; 36: 824-830.

Summers C, Hammons AL, Mitchell LE, Woodside JV, Young IS, Yarnell JWG, Evans A, Whitehead AS. Influence of the cystathionine β-synthase 844ins68 and 5,10-methylene tetrahydrofolate reductase 677C>T polymorphisms on folate and homocysteine concentrations. *Eur J Hum Genet*, 2008; 16: 1010-1013.

Penning, TM, Lerman, C: Genomics of smoking exposure and cessation: lessons for cancer prevention and treatment. *Cancer Prevention Res*. 2008; 1: 80-83

Huang Y, Khartulyari S, Morales M, Stanislawska-Sachadyn A, Von Feldt JM, Whitehead AS, Blair IA. Quantification of key red blood cell folates from subjects with defined MTHFR

677C>T genotypes using stable isotope dilution liquid chromatography/mass spectrometry. *Rapid Commun Mass Spectrom*, 2008; 22: 2403-2412.

Lu Z-Y, Morales M, Khartulyari S, Mei M, Murphy KM, Stanislawska-Sachadyn A, Summers CM, Huang Y, Von Feldt JM, Blair IA, Mitchell L, Whitehead AS. Genetic and biochemical determinants of serum concentrations of monocyte chemoattractant protein-1, a potential neural tube defect risk factor. *Birth Defects Research (PartA) Clin Mol Teratol*, 2008; 82: 736-741.

Lu Z-Y, Jensen LE, Huang Y, Kealey C, Blair IA, Whitehead AS. The up-regulation of monocyte chemoattractant protein-1 (MCP-1) in endothelial cells maintained in low folate conditions is mediated by the p38 MAPK pathway. *Atherosclerosis*, 2009; 205: 48-54.

Olson KC, Dellinger RW, Zhong Q, Sun D, Spratt TE, Lazarus P. Functional Characterization of Low-Prevalence Missense Polymorphisms in the UGT1A9 Gene. *Drug Metabol Dispos*. 2009; 37: 1999-2007.

Quinn, AM, Harvey, RG, Penning TM. Oxidation of PAH *trans*-dihydrodiols by human aldo-keto reductase AKR1B10. *Chem Res Toxicol*. 2008; 21: 2207-2215.

Stanislawska-Sachadyn A, Mitchell LE, Woodside JV, Buckley P, Kealey C, Young IS, Scott JM, Murray L, Boreham CA, McNulty H, Strain JJ, Whitehead AS. The reduced folate carrier (*RFC1*) 80G>A polymorphism is associated with red cell folate concentrations among women. *Ann Hum Genet*, 2009; 73: 484-491.

Mitchell LE, Morales M, Khartulyari S, Huang Y, Murphy KM, Mei M, Von Feldt JM, Blair IA, Whitehead AS. Folate and homocysteine phenotypes: comparative findings using research and clinical laboratory data. *Clin Biochem*, 2009; 42: 1275-1281.

Hammons AL, Summers C, Lu Z-Y, Woodside JV, Young IS, Scott JM, Murray L, Boreham CA, McNulty H, Strain JJ, Mitchell LE, Whitehead AS. Folate/homocysteine phenotype and *MTHFR* 677C>T genotype are determinants of MCP-1 concentrations in young women. *Clin Immunol*, 2009; 133: 132-137.

Balliet RM, Chen G, Gallagher CJ, Dellinger RW, Sun D, Lazarus P. Characterization of UGTs active against suberoylanilide hydroxamic acid (SAHA) and association between SAHA glucuronidation activity phenotype with UGT genotype. *Cancer Res* 2009; 69: 2981-2989.

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.

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.* 2010; 38: 484-90.

Subsequent to May 2010

Chen G, Giambrone NE Jr, Dluzen DF, Muscat JE, Berg A, Gallagher CJ, Lazarus P. Glucuronidation genotypes and nicotine metabolic phenotypes: importance of functional UGT2B10 and UGT2B17 polymorphisms. *Cancer Res.* 2010; 70: 7543-52.

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.

Meireles SI, Esteves GH, Hirata R Jr, Peri S, Devarajan K, Slifker M, Mosier SL, Peng J, Vadhanam MV, Hurst HE, Neves EJ, Reis LF, Gairola CG, Gupta RC, Clapper ML. Early changes in gene expression induced by tobacco smoke: Evidence for the importance of estrogen within lung tissue. Cancer Prev Res (Phila). 2010 3:707-17. Epub 2010 Jun 1.

Gallagher CJ, Balliet RM, Sun D, Chen G, Lazarus P.Sex differences in UDP-glucuronosyltransferase 2B17 expression and activity. *Drug Metab Dispos.* 2010; 38: 2204-9.

Bushey RT, Chen G, Blevins-Primeau AS, Krzeminski J, Amin S, Lazarus P. Characterization of UDP-glucuronosyltransferase 2A1 (UGT2A1) variants and their potential role in tobacco carcinogenesis. Pharmacogen Genomics 2011; 21:55-65.

Jones NR, Spratt TE, Berg AS, Muscat JE, Lazarus P, Gallagher CJ. Association studies of excision repair cross-complementation group 1 (*ERCC1*) haplotypes with lung and head and neck cancer risk in a Caucasian population. *Cancer Epidemiol.* 2011; 35: 175-81. Epub 2010 Sep 21.

Shatalova EG, Klein-Szanto AJ, Devarajan K, Cukierman E, Clapper ML Estrogen and cytochrome P450 1B1 contribute to both early- and late-stage head and neck carcinogenesis. Cancer Prev Res (Phila). 2011; 4 107-15.

Zhang L, Jin Y., Chen M, Huang, Harvey RG, Blair IA, Penning TM: Detoxication of structurally diverse polycyclic aromatic hydrocarbon (PAH) *o*-quinones by human recombinant catechol-*O*-methyltransferase (COMT) via *O*-methylation of PAH catechols**.** *J. Biol. Chem.* 2010; 286: 25644-54.

Rangiah K, Hwang WT, Mesaros C, Vachani A, Blair IA.: Nicotine exposure and metabolizer phenotypes from analysis of urinary nicotine and its 15 metabolites by LC-MS. *Bioanalysis.* 2011; 3: 745-61. doi: 10.4155/BIO.11.42.

Olson KC, Sun D, Chen G, Sharma AK, Amin S, Ropson IJ, Spratt TE, Lazarus P. Characterization of dibenzo[*a,l*]pyrene-*trans*-11,12-diol (Dibenzo[*def,p*]chrysene) glucuronidation by UDP-glucuronosyltransferases. *Chem Res Toxicol.* 2011; 24: 1549-59

Crampsie MA, Jones NR, Das A, Aliaga C, Desai D, Lazarus P, Amin S, Sharma AK. Phenylbutyl isoselenocyanate modulates Phase I and II enzymes and inhibits 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone induced DNA adducts in mice. Cancer Prev Res 2011 Jul 27.

Shultz CA, Quinn AM, Park JH, Harvey RG, Bolton JL, Maser E, Penning TM.: Specificity of human aldo-keto reductases, NAD(P)H:Quinone oxidoreductase and carbonyl reductases to redox cycle polycyclic aromatic hydrocarbon diones and 4-hydroxyequilenin-*o*-quinone. *Chem Res Toxicol.* 2011 Sep 12. [Epub ahead of print]

Lu D, Harvey RG, Blair, IA, and Penning, TM: Quantitation of the benzo[*a*]pyrene metabolome in human bronchoalveolar (H358) cells by stable isotope dilution liquid chromatography-atmospheric chemical ionization mass spectrometry. *Chem. Res. Toxicol.* In revision.

Tables and Figures

Table 1. University of Pennsylvania (Cumulative through 02-28-11) Combined Penn and Temple		
	Lung Cancer Cases	**Controls**
Gender		
Male	326	208
Female	258	194
Race		
White	466	255
African-American	105	139
Asian	9	6
Other/Unknown	4	2
TOTAL	**584**	**402**

Table 2. Fox Chase Cancer Center (Cumulative through 02-28-11)		
	Lung Cancer Cases	**Controls**
Gender		
Male	75	25
Female	73	74
Race		
White	126	86
African-American	17	5
Asian	1	4
Other/Unknown	4	4
TOTAL	**148**	**99**

Table 3. Combined Archival African American Samples		
	Lung Cancer Cases	**Controls**
Gender		
Male	60	43
Female	61	35
Race		
White	0	0
African-American	121	78
Asian	0	0
Other/Unknown	0	0
TOTAL	**121**	**78**

Table 4. Penn State Hershey Medical Center (Cumulative through 02-28-11) Combined Hershey and Pinnacle		
	Lung Cancer Cases	**Controls**
Gender		
Male	159	171
Female	119	191
Race		
White	264	310
African-American	12	42
Asian	2	0
Other/Unknown	0	10
TOTAL	**278**	**362**

Table 5. Biomarker Analysis Performed

Biomarker analysis	Penn samples		Hershey samples		Totals
	Cases	Controls	Cases	Controls	
8-oxo-dGuo in urine*	280	260	103	103	746
Nicotine and 15 metabolites in urine*	280	260	103	103	746
Total NNAL in urine*	280	260	103	103	746
Total 1-OH-pyrene in urine*	280	260	103	103	746
Creatinine in urine	280	260	103	103	746
5-MTHF in RBC**	200	200	0	0	400
THF in RBC**	200	200	0	0	400
5,10-METHF RBC**	200	200	0	0	400
Folic acid in RBC**	200	200	0	0	400
5-MTHF in plasma**	200	200	0	0	400
THF in plasma**	200	200	0	0	400
5,10-METHF in plasma**	200	200	0	0	400
Folic acid in plasma**	200	200	0	0	400
Hematocrit	200	200	0	0	400
Total biomarker samples analyzed					**7330**

*Urinary biomarkers are reported as pmol/mL or nmoles/mmol creatinine
**RBC and plasma folates are reported as concentration (nM)
NNAL=4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol
5-MTHF = 5-methyltetrahydrofolate
THF = tetrahydrofolate
5,10-METHF = 5,10-methenyltetrahydrofolate

Table 6: SNPs with p-value <.001. Minor Allele frequencies in case and control are also reported.

rs#	chr	p-value	gene	MAF in case	MAF in control
rs743535	10	0.000436	CYP2E1	0.051334702	0.090609555
rs7092584	10	0.000723	CYP2E1	0.057494867	0.101317957
rs11593189	10	0.000795	SYCE1/CYP2E1	0.058521561	0.097359736
rs8192772	10	0.000911	CYP2E1	0.034907598	0.070840198

Table 7: Analysis of the MTHFR677 C>T Variant

GENE	rs	chr	P values	Allele frequency in case	Allele frequency in control
MTHFR 677	rs1801133	1	0.2052	T: .366 C: .634 Counts: TT: 68 CT: 220 CC: 198	T: .389 C: .611 Counts: TT: 100 CT: 272 CC: 235

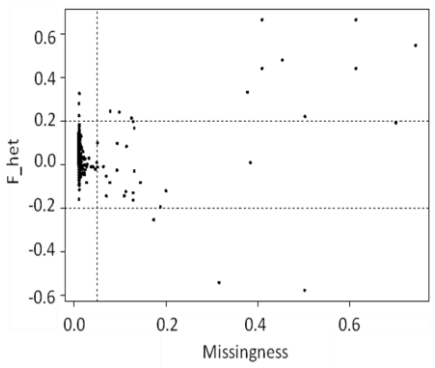

Figure 1. Fraction of genotype missingness and heterozygosity

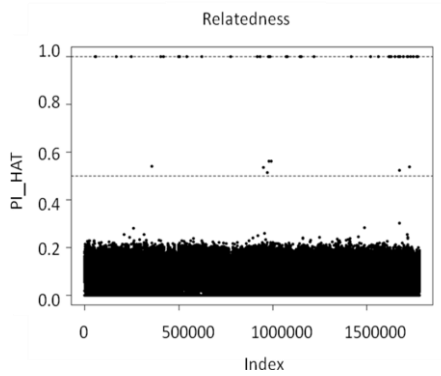

Figure 2. IBD analysis to identify duplicates and related individuals

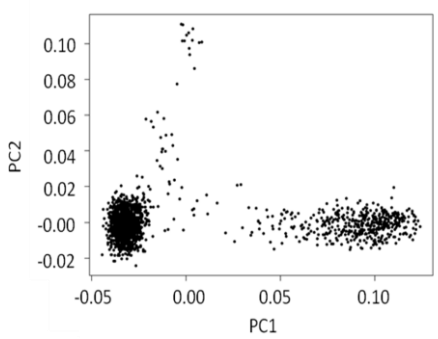

Figure 3. MDS analysis to identify inferred Caucasians and African Americans

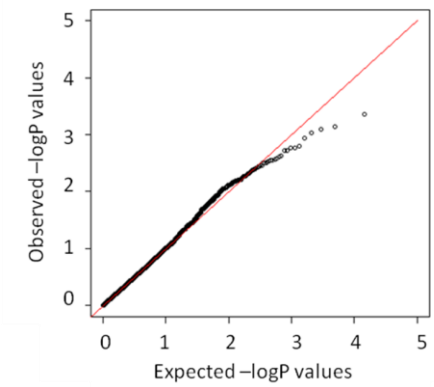

Figure 5. Q-Q plot for genetic Association analysis in Caucasians (n=1094)

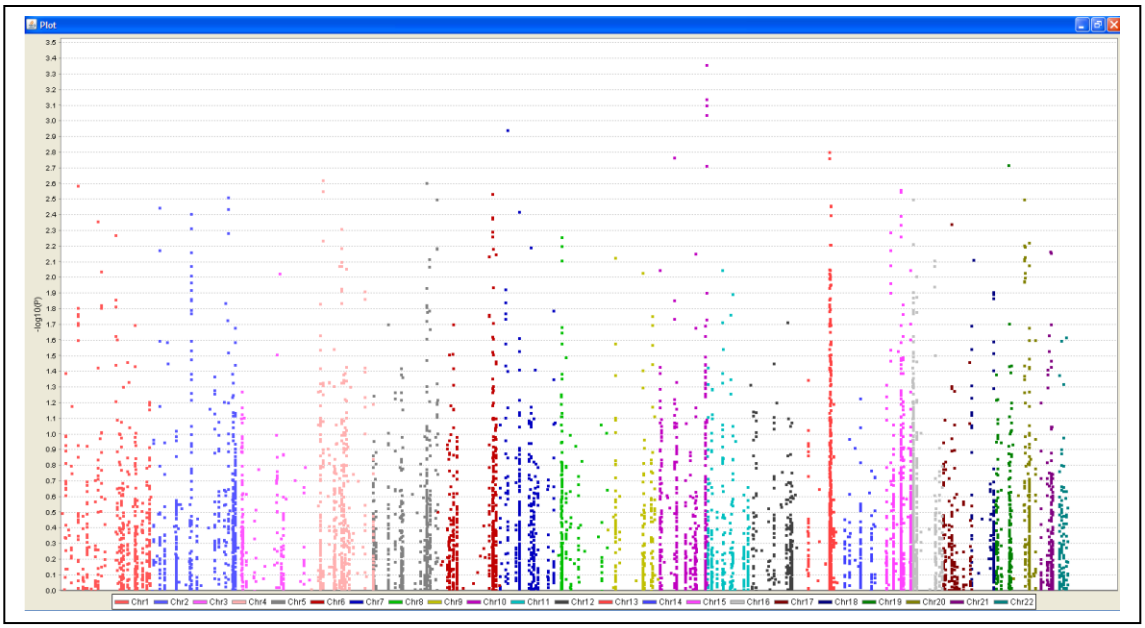

Figure 4. Manhattan plot for genetic association analysis in Caucasians (n =1094)