# Final Progress Report for Research Projects Funded by Health Research Grants

Instructions: Please complete all of the items as instructed. Do not delete instructions. Do not leave any items blank; responses must be provided for all items. If your response to an item is "None", please specify "None" as your response. "Not applicable" is not an acceptable response for any of the items. There is no limit to the length of your response to any question. Responses should be single-spaced, no smaller than 12-point type. The report **must be completed using MS Word**. Submitted reports must be Word documents; they should not be converted to pdf format. Questions? Contact Health Research Program staff at 717-783-2548.

1. **Grantee Institution:** The Wistar Institute

2. **Reporting Period (start and end date of grant award period):** 01/01/09 – 06/30/10

3. **Grant Contact Person (First Name, M.I., Last Name, Degrees):** Russel E. Kaufman, M.D.

4. **Grant Contact Person's Telephone Number:** 215-898-3926

5. **Grant SAP Number:** SAP #4100047657

6. **Project Number and Title of Research Project:** Project 4: Diagnosis of Lung Cancer from Peripheral Blood Using Genomics

7. **Start and End Date of Research Project:** 01/01/09 – 06/30/10

8. **Name of Principal Investigator for the Research Project:** Louise C. Showe, Ph.D.

9. **Research Project Expenses**.

   9(A) Please provide the amount of health research grant funds spent on this project for the entire duration of the grant, including any interest earned that was spent:

   $ 401,938.16

   9(B) Provide the last names (include first initial if multiple individuals with the same last name are listed) of **all** persons who worked on this research project and were supported with health research funds. Include position titles (Principal Investigator, Graduate Assistant, Post-doctoral Fellow, etc.), percent of effort on project and total health research funds expended for the position. For multiple year projects, if percent of effort varied from year to year, report in the % of Effort column the effort by year 1, 2, 3, etc. of the project (x% Yr 1; z% Yr 2-3).

| Last Name | Position Title | % of Effort on Project | Cost (Salary & FB) |
|---|---|---|---|
| Showe, L. | Principal Investigator | 56% | $ 116,539 |
| Showe, M. | Sr. Staff Scientist | 45% | $ 61,654 |
| Nikonova | Staff Scientist | 70% - 2 months | $ 5,564 |
| Kossenkov | Postdoctoral Fellow | 28% | $ 23,633 |
| Billouin | Research Assistant | 5% - 6 months | $ 1,311 |
| | | | |
| | | | |

9(C) Provide the names of **all** persons who worked on this research project, but who *were not* supported with health research funds. Include position titles (Research Assistant, Administrative Assistant, etc.) and percent of effort on project. For multiple year projects, if percent of effort varied from year to year, report in the % of Effort column the effort by year 1, 2, 3, etc. of the project (x% Yr 1; z% Yr 2-3).

| Last Name | Position Title | % of Effort on Project |
|---|---|---|
| Calen Nichols | Research Assistant | 10% |
| Celia Chang | Staff Scientist | 10% |
| Wen Hwai Horng | Research Assistant | 5% |
| | | |
| | | |
| | | |
| | | |

9(D) Provide a list of **all** scientific equipment purchased as part of this research grant, a short description of the value (benefit) derived by the institution from this equipment, and the cost of the equipment.

| Type of Scientific Equipment | Value Derived | Cost |
|---|---|---|
| None | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |

**10. Co-funding of Research Project during Health Research Grant Award Period.** Did this research project receive funding from any other source during the project period when it was supported by the health research grant?

Yes____X _____ No _____

If yes, please indicate the source and amount of other funds: The Wistar Institute $30,416

**11. Leveraging of Additional Funds**

11(A) As a result of the health research funds provided for this research project, were you able to apply for and/or obtain funding from other sources to continue or expand the research?

Yes____X____ No_____

If yes, please list the applications submitted (column A), the funding agency (National Institutes of Health—NIH, or other source in column B), the month and year when the application was submitted (column C), and the amount of funds requested (column D). If you have received a notice that the grant will be funded, please indicate the amount of funds to be awarded (column E). If the grant was not funded, insert "not funded" in column E.

Do not include funding from your own institution or from CURE (tobacco settlement funds). Do not include grants submitted prior to the start date of the grant as shown in Question 2. If you list grants submitted within 1-6 months of the start date of this grant, add a statement below the table indicating how the data/results from this project were used to secure that grant.

| A. Title of research project on grant application | B. Funding agency (check those that apply) | C. Month and Year Submitted | D. Amount of funds requested: | E. Amount of funds to be awarded: |
|---|---|---|---|---|
| Signatures for diagnosis & prognosis of NSCLC from gene expression in blood | ☒NIH<br>☐ Other federal (specify:_____)<br>☐ Nonfederal source (specify: _____) | 07/2010 | $3,211,373 Direct +indirect 5years | $ TBD |
| Clinical Test for Lung Cancer Using Whole Blood Samples | ☐NIH<br>☐ Other federal (specify:_____)<br>☒Nonfederal source (Science Center QED Program) | 08/2010 | $100,000 | $ TBD |
| | ☐NIH<br>☐ Other federal (specify:_____ _____)<br>☐ Nonfederal source (specify: _____) | | $ | $ |

11(B) Are you planning to apply for additional funding in the future to continue or expand the research?

Yes ____X_____ No_____

If yes, please describe your plans:

We have submitted a grant to NCI and another to the Philadelphia Science Center QED Program to continue this work. The grants are listed above. We are also in contact with two companies for sponsored research support.

**12. Future of Research Project.** What are the future plans for this research project?

We are now developing and testing new methods of sample collection that will be necessary to allow us to move this project to a commercially viable state. These new techniques will allow samples for testing to be collected and tested without any special processing to isolate particular types of blood cells.

**13. New Investigator Training and Development**. Did students participate in project supported internships or graduate or post-graduate training for at least one semester or one summer?

Yes _____X_____ No_____

If yes, how many students? Please specify in the tables below:

| | Undergraduate | Masters | Pre-doc | Post-doc |
|---|---|---|---|---|
| Male | | | | |
| Female | | 2 | | |
| Unknown | | | | |
| **Total** | | **2** | | |

| | Undergraduate | Masters | Pre-doc | Post-doc |
|---|---|---|---|---|
| Hispanic | | 1 | | |
| Non-Hispanic | | 1 | | |
| Unknown | | | | |
| **Total** | | **2** | | |

| | Undergraduate | Masters | Pre-doc | Post-doc |
|---|---|---|---|---|
| White | | | | |
| Black | | 1 | | |
| Asian | | 1 | | |
| Other | | | | |
| Unknown | | | | |
| **Total** | | **2** | | |

**14. Recruitment of Out-of–State Researchers**. Did you bring researchers into Pennsylvania to carry out this research project?

Yes_ ____ No ____X____

If yes, please list the name and degree of each researcher and his/her previous affiliation:

**15. Impact on Research Capacity and Quality**. Did the health research project enhance the quality and/or capacity of research at your institution?

Yes ____X_____ No_____

If yes, describe how improvements in infrastructure, the addition of new investigators, and other resources have led to more and better research.

This project allowed us to develop new approaches to identify gene expression information from clinical samples with diagnostic and prognostic utility. In particular, application of the Bioinformatics approaches that we developed for our own work has led to new collaborations with other laboratories interested in applying our approaches to their own samples of interest.

**16. Collaboration, business and community involvement.**

16(A) Did the health research funds lead to collaboration with research partners outside of your institution (e.g., entire university, entire hospital system)?

Yes ___X____ No_____

If yes, please describe the collaborations:

We have established collaborations at the University of Pennsylvania, NYU Medical Center, and Brigham and Women's Hospital.

16(B) Did the research project result in commercial development of any research products?

Yes_____ No ____X_____

If yes, please describe commercial development activities that resulted from the research project:

16(C) Did the research lead to new involvement with the community?

Yes____X____ No_____

If yes, please describe involvement with community groups that resulted from the research project:

Because of my research in the lung cancer field, I was appointed to the Board of Directors of the PA Cancer Control Consortium which addresses disparities in cancer treatment across the state.

**17. Progress in Achieving Research Goals, Objectives and Aims.**

List the project goals, objectives and specific aims (as contained in the grant application's strategic plan). Summarize the progress made in achieving these goals, objectives and aims for the entire grant award period. Indicate whether or not each goal/objective/aim was achieved; if something was not achieved, note the reasons why. Describe the methods used. If changes were made to the research goals/objectives/aims, methods, design or timeline since the original grant application was submitted, please describe the changes. Provide detailed results of the project. Include evidence of the data that was generated and analyzed, and provide tables, graphs, and figures of the data. List published abstracts, poster presentations and scientific meeting presentations at the end of the summary of progress; peer-reviewed publications should be listed under item 20.

This response should be a DETAILED report of the methods and findings. It is not sufficient to state that the work was completed. Insufficient information may result in an unfavorable performance review, which may jeopardize future funding. If research findings are pending publication you must still include enough detail for the expert peer reviewers to evaluate the progress during the course of the project.

Health research grants funded under the Tobacco Settlement Act will be evaluated via a performance review by an expert panel of researchers and clinicians who will assess project work using this Final Progress Report, all project Annual Reports and the project's strategic plan. After the final performance review of each project is complete, approximately 12-16 months after the end of the grant, this Final Progress Report, as well as the Final Performance Review Report containing the comments of the expert review panel, and the grantee's written response to the Final Performance Review Report, will be posted on the CURE Web site.

**There is no limit to the length of your response. Responses must be single-spaced below, no smaller than 12-point type. If you cut and paste text from a publication, be sure symbols print properly, e.g., the Greek symbol for alpha (α) and beta (ß) should not print as boxes (□) and include the appropriate citation(s). DO NOT DELETE THESE INSTRUCTIONS.**

*Overview*

Non-small cell lung cancer (NSCLC) is the most common cause of premature death from malignant disease in western countries. Early detection followed by surgery remains the most effective treatment. The diagnosis of patients who have lung nodules of unknown etiology identified on screening chest X-rays or computerized axial tomography (CT) scans is a common and important clinical problem for internists, surgeons, and pulmonologists. A blood test that

could more accurately determine the risk of malignant disease in patients with lung nodules would be extremely valuable and have very important economic implications by reducing unnecessary surgery, biopsies, Positron Emission Tomography (PET) scans, and/or repeated CT scans. *We hypothesize that a specific and diagnostic gene expression profile for NSCLC can be identified in the peripheral blood mononuclear cells (PBMC) of lung cancer patients*. Preliminary studies have shown clear differences in gene expression profiles among patients with early stage adenocarcinomas of the lung (AC) and the appropriate high risk control group of smokers and ex-smokers matched by age, gender, and race. The goal of this project was to expand these preliminary studies to validate our preliminary results using a larger study population in a statistically rigorous manner.

*Summary of Research Completed*

<u>Aim 1. Using the Illumina arrays, we will optimize our diagnostic signatures for early stage lung cancer.</u>

1a. *Assess the accuracy of a global diagnostic signature to identify patients with NSCLCs of various cell types, and compare the accuracy of a more global diagnostic to cell type specific diagnostic signatures.*

Our previous studies comparing gene expression in 137 NSCLC patients and 91 non-healthy controls (NHC) with non-malignant smoking related disease identified a 29 gene expression signature with 91% sensitivity and 80% specificity **(Figure 1)**. We have now assessed the accuracy of that signature to identify sub-classes of patients and control subjects. Those results are shown in **Table 1**. It should be noted that Lung Squamous Cell Carcinomas (LSCC) were classified with higher accuracy than the adenocarcinomas (AC) and that accuracy increased incrementally with advanced stage. Also, it should be noted that controls with COPD were more accurately identified than patients with histologically confirmed benign nodules.

We also assessed accuracy based on smoking status as shown in **Table 2**. The lowest accuracy determined was for NHCs that were current smokers, but this was such a small class that the result did not reach significance. It should be noted that the 9 NSCLCs in the **"***never smokers***"** category were classified with 89% accuracy, suggesting that, although lung cancer in "never smokers" differs in many ways from that in smokers, it is still accurately identified by the changes in the immune profile that we are monitoring. The results associated with Aim1a have been published in *Cancer Research, Showe et al 2009.*

We also determined the positive predictive value and the negative predictive value for the 29-gene NSCLC classifier, and the results in comparison to the Spira et al 2007 data are shown in **Table 3**.

1b. *Determine if accuracy can be improved by developing specific signatures that take into account cancer subtype and stage and in particular develop high accuracy in distinguishing benign from malignant nodules.*

Gene Expression Classifier for Lung Nodules:
The 29-gene signature was generated with control subjects from two different at-risk populations. About half (50) were "high risk" based on underlying lung disease and smoking history, while an additional 41 had been further diagnosed by CT or chest X-ray with lung nodules and were to undergo surgical evaluation. When we calculate classification accuracy for the two control populations separately, the 29-gene classifier has a specificity of 89% if only the controls without lung nodules are considered, whereas the specificity is 71% for controls with confirmed benign nodules (**Table 1**). To refine the classification, we carried out the following analysis. Since the nodule patients are from the same population as those with confirmed malignant tumors, we applied SVM-RFE (see *Methods*, below) with cross-validation using only the data from the 41 patients with benign nodule as the control group and data from a randomly selected group of 54 case samples with COPD. All of the nodule samples were obtained from a population of patients with lung nodules or masses undergoing surgical evaluation. The resulting 24 gene "nodule" classifier **(Table 4)** had an overall accuracy of 79%, and the specificity for the benign nodule controls is 80% (compared to 71% for the nodules in the analysis with the mixed control group). We expect that, with an increased nodule study population, the accuracy in identifying this important class of cases will be increased.

Taken together, these results suggest that it is possible to develop a non-invasive test for the presence of lung cancer using gene expression in peripheral blood samples. This test may be used in conjunction with present standards for detection and may help to reduce the large numbers of false positive results presently resulting from CT scans.

1c. Evaluate *changes in diagnostic signatures present in pre-surgery and post therapy (after curative surgery) samples to identify signatures that may be used to assess the probability of recurrence in a long term goal.*

Effect of Tumor Removal on Individual Classification Scores
To further assess whether the signature we have detected for lung cancer was truly due to the presence of the tumor, we asked whether the SVM scores using the 29 gene classifier that we had identified was altered after removal of the tumor. Eighteen of the NSCLC patients not included in the selection of the 29 gene classifier had post-resection blood samples that were collected 2-5 months after and before any chemotherapy. To assess how the removal of the tumor affected the *29-gene SVM* scores calculated for the pre-surgery samples, we also determined the SVM scores for the *post*-resection samples from each pair **(Figure 2).** Of the 14 patients that classified as cancer in the validation set (i.e., had positive SVM scores), 12 (86%) showed significant decreases in their SVM scores in the post-resection samples. Five of these post-surgery samples (4, 5, 6, 10, and 13) had clearly negative SVM scores and would be classified as non-cancer samples in the analysis. The post-surgery changes in two of the pairs, numbers 8 and 16, were not statistically significant. Of the 4 misclassified, pre-surgery patients, 1 showed a highly decreased score, 1 was not significantly changed and 2 showed increases. Although the time intervals between the first and second samples ranged between 2 and 5 months, there was no obvious relationship between the change in the scores and the time to post-resection sample collection. In the large majority of the patients (13 out of 18), tumor removal was associated with a decrease in the cancer signature. Additional pre-post sample pairs are continuing to be analyzed. Additional preliminary results of the pre-post surgery analyses are presented after the Aim 2 results.

Aim 2. *Validate the generalizability of our signatures on subsets of case and control samples from external validation sites not used to develop the diagnostic signature.*

Validation of the 29-Gene Classifier on Independent Samples

Although we had used cross-validation to establish our 29-gene classifier, to further validate the utility of the classifier for analyzing new samples, we assessed the classification accuracy using new samples not included in the 29-gene selection process. The validation set included 38 NSCLC samples and 17 controls. Twenty-seven of the validation samples were collected at the New York University Medical Center (NYUMC) Lung Cancer Biomarker Center, an Early Detection Research Network (EDRN) Clinical and Epidemiologic Validation Center. The dataset included 12 Stage 1 NSCLC (5 of whom were never smokers) and 15 smoker and ex-smoker controls. Six of the controls were diagnosed by serial CT scans as having non-malignant Ground Glass Opacities (GGO). No GGO patient samples were included in our original training set. The RNA for these samples was prepared at NYUMC. An additional 26 patients and 2 control samples were collected at the University of Pennsylvania Medical Center (Penn) and had not been analyzed previously. The classification algorithm was applied to these samples with no knowledge of whether a sample is a case or control (see *Methods*). The classification for the validation set is shown in **Figure 3.** The overall accuracy for the validation set was 78%, with 76% sensitivity and 82% specificity. This small decrease in accuracy and sensitivity (although with an increase in specificity) was not unexpected since the samples were not specifically collected for these studies; and, as a result, the sample collection and RNA purification were not standardized for the NYUMC samples. Collection of samples from the NYUMC and Penn sites are continuing.

Effect of Tumor Removal on Individual Classification Scores

*Additional pre-post studies (Kossenkov et al 2010 paper in preparation)*: During the past funding period, we have further analyzed the differences in gene expression in the peripheral immune system taken before and after removal of NSCLC. These studies have begun to provide information that suggests the importance of pursuing these studies in a much larger sample size. To further assess the differences between pre and post surgery samples, we directly compared these two classes of samples. Because of the unique relationship of our cancer and control cases, expression of only 4 genes are required for 100% accurate discrimination of the two classes. A comparison of the post-surgery samples and patients with non-malignant lung disease, with and without the presence of benign nodules, indicates that they form a separate class in a hierarchy of relationships. Even though all four classes of patients share many characteristics of lung disease, their relative relationships to NSCLCs based on gene expression are: NSCLC>benign nodules>post surgery samples>samples with other lung diseases. We find significant alterations post-surgery in pathways associated with the innate immune response. Altered pathways associated with the analyses in Aim 1 and the overlapping pathways in the pre-post surgery analysis are shown in **Figure 4**.

*Milestones*

Aim 1: Continue to collect samples from Stage I+II non small cell lung cancers and process on microarrays – collect and analyze data.

- Increase collection of samples from pre and post-surgery benign nodules .
- Collect post-surgery NSCLC samples.

Aim 2: Collect samples from different locations and process on microarrays – collect data.

We continue to collect samples for both cases and controls from Penn and NYUMC. We identified a source of pre/post surgery samples for NSCLCs and benign nodules at Brigham and Womens Hospital and wrote a grant proposal to NCI to process 500 of these samples in the next five years. In addition, we have collected approximately 400 samples from NYUMC at this time, and approximately 300 have been processed on arrays for analysis.

Measures Of Success

We published a manuscript of the study of the 228 PBMC samples in *Cancer Research* in December of 2009. We have applied for a patent based on the technology and have had a marketing assessment done of the clinical utility. We are negotiating with a number of biotech companies, including Quest and Phoenix Ltd. Two additional manuscripts are in preparation.

We analyzed miRNA transcription profiles that differed in pre vs. post-surgery samples. We found expression of five miRNAs to be universally *upregulated* in pre- compared to the post-surgery samples, and then computationally identified targets of those miRNAs within the differentially expressed genes *downregulated* in pre- compared to post-surgery samples. miRNA targets were significantly enriched for transcription factors, including several key factors in immune regulation.

We have now established a collaboration with a lung cancer surgeon at Brigham and Womens Hospital in Boston. He is now carefully collecting samples in PAXgene tubes pre- and post-surgery for both NSCLCs as well as for benign nodules. We have applied for NCI funding to pursue these studies. In addition to gene expression, we will also carry out microRNA expression studies with the intent to develop a gene expression signature which includes markers for both coding and non-coding RNAs which may provide a more robust classifier.

Validation of the 29-Gene Classifier on Independent Samples

Validation studies are continuing with samples being collected at various locations. Because of the large differences in the quality of the PBMC samples acquired from several different locations, we are continuing our studies by analyzing samples in whole blood collected in the more easily standardized PAXgene system. We have demonstrated the similarities in gene expression in PAXgene and immediately processed RNA (**Figure 2**). PAXgene tubes require the collection of only 2.5 mls of blood, which is immediately stabilized by a reagent in the collection tube. As it does not require sample processing in a timely manner at the site of sample collection, this approach facilitates highly controlled sample collection at sites that lack equipment or expertise to process samples at the time of collection. This approach has greatly reduced the burden on the collection and has increased our ability to collect many more samples. We have now received *~400* PAXgene samples from NYUMC and 50 samples from UPENN, and have also established collaborations with investigators at Brigham and Womens Hospital who are collecting pre-post surgery samples from patients and benign nodule controls in PAXgene. A grant application to study these samples has been submitted to NCI. We are presently processing

the samples from NYUMC and Penn, and an MTA for accessing the samples from Brigham and Womens Hospital is underway.

*Methods*

Study Populations**.** Study participants (Showe et al 2009) for the initial training sets were recruited from Penn during the period 2003 through 2007 – 91 subjects with a history of tobacco use without lung cancer, including 41 subjects that had one non-calcified lung nodule diagnosed as benign after biopsy, and 137 patients with newly diagnosed, histopathologically confirmed, non-small-cell lung cancer. All participants had blood collection in conjunction with a clinical visit or just prior to surgery. None of the case subjects had received any cancer therapy prior to blood collection. Samples in an independent validation set were from 27 cases and controls recruited at NYUMC and 28 new patients recruited at Penn, all under IRB approval. PAXgene samples from Penn are from a similar demographic as the PBMC. The PAXgene samples from NYU are collected as part of a CT screening program of smokers and ex-smokers. Although the blood collection system has been altered, the goals of the proposal remain the same. The decision to make this change in sample collection was the realization that sample collection had to be simplified for commercial application.

PBMC Collection and Processing: Blood samples were drawn in two "CPT" tubes (BD). PBMC were isolated within 90 minutes of blood draw, washed in PBS, transferred into RNAlater (Ambion), and stored at 4$^{o}$C overnight before transfer to -80 $^{o}$C. A subset of patient PBMCs was analyzed by flow cytometry, with anti-CD3, -CD4, -CD8, -CD14, -CD16, -CD19, or -CD56 antibodies or isotype controls (BD Biosciences), and analyzed using FlowJo software. Samples from NYUMC were processed within 2 hours from collection; PBMC were transferred to Trizol (Invitrogen) and stored at -80 $^{o}$C. Extracted RNA was transferred to The Wistar Institute for further processing. RNA purification of the Penn samples was carried out at Wistar using TriReagent (Molecular Research), as recommended and controlled for quality using the Bioanalyzer. Only samples with 28S/16S ratios >0.75 were used for further studies. The NYUMC samples were DNAse-treated before hybridization. Samples were processed as mixed batches of cases and controls and hybridized to the Illumina WG-6v2 human whole genome bead arrays.

Analysis**:** Classification was performed using a Support Vector Machine with recursive feature elimination (SVM-RFE) using random, tenfold, cross-validation repeated 10 times. Classification scores for each tested sample were recorded at each reduction step, down to a single gene. Average accuracy for each reduction step was calculated and all the genes at the points of maximal accuracy formed the initial discriminator, which then underwent additional reduction to form the final discriminator. Significance of the changes in the SVM score before and after surgery was determined with a one-sided t-test.

Validation of the Classifier on Independent Samples**:** Each of the genes in the signature from SVM analysis of the microarray data identified in the training set is assigned a coefficient that defines its importance in the classifier. In validating or testing the accuracy of the signature on new samples that are not identified by class association, the analysis is carried out essentially as follows:

The signature is applied as an equation of the form:

$$X = a[A] + b[B] + c[C] \ldots + z[Z] + constant$$

where A, B, C, etc., are the microarray expression levels of each of the signature genes, and a, b, c, etc., are the coefficients by which each expression level is multiplied to give a value for X (the classification score). The expression levels of the 29 genes [A, B, C…Z] determined by microarray for a new patient are each multiplied by the appropriate coefficient (a, b, c…z) to determine a classification score, "X." If the threshold value of X is set to zero, then patients with positive scores will be declared to have malignant disease and those with negative scores will be called non-malignant. The higher the positive score, the greater is the confidence of malignancy; and the more negative the score, the greater is the confidence of no malignancy.

*Milestone(s)*

The original milestones have all been accomplished. We continue to collect samples for both cases and controls from Penn and NYUMC, and we have switched the sample collection to the PAXgene system as a more viable approach for development for clinical use. During the course of this project, we have collected approximately 400 samples from NYUMC to date and approximately 300 have been processed on arrays for analysis. We have identified a source of pre/post surgery samples for NSCLCs and benign nodules in PAXgene at Brigham and Womens Hospital. We have written a grant proposal to NCI to process 500 samples in the next five years.

**Measures Of Success**

We published a manuscript of the study of the 228 PBMC samples in *Cancer Research* in December 2009. An additional manuscript describing the reanalysis of the data in the 2009 publication with patient outcome data now collected is in preparation (Vachani et al, in preparation). We will also be submitting a third manuscript that focuses on the further analysis of differences between the pre and post-surgery gene expression profiles and what they can tell us about the interactions of the tumor and the peripheral immune system.

A patent application, PCT/US2008/013450 ("Method for Diagnosing Lung Cancers using Gene Expression Profiles in Peripheral Blood Mononuclear Cells"), was filed 12/05/2008. National phase applications have been filed in the US and Europe. This patent application, jointly owned by The Wistar Institute and the University of Pennsylvania, claims compositions for and methods of diagnosing and staging lung cancer using purified white blood cells (PBMC) and includes multiple claims of sets of genes used in the diagnosis and/or staging of the cancer. The European patent application was filed at the request of a local venture firm that is interested in commercializing this test. Our application to the QED program at the Philadelphia Science Center has passed the first level of review.

## Figures and Tables

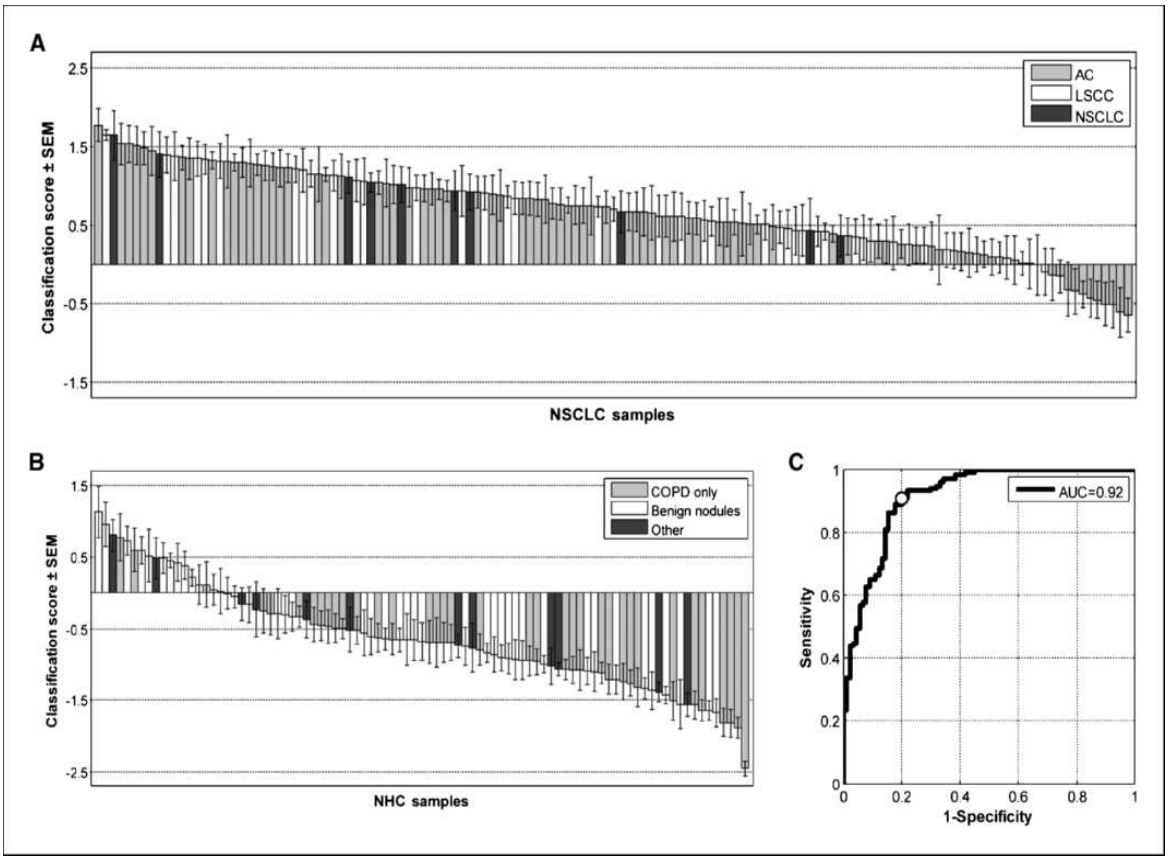


**Figure 1**. Classification scores assigned by the NSCLC classifier to 137 NSCLC patients and 91 patients with nonmalignant lung disease. A positive score indicates classification as a cancer; a negative score as a nonmalignant disease. The column heights are a measure of how well the sample is classified by the SVM algorithm for the 29 genes, and the error bars are a measure of the classification variance across the 100 resamplings. A, NSCLC patients: AC, adenocarcinoma; LSCC, lung squamous cell carcinoma; NSCLC, samples not further characterized. B, NHCs include patients with nonmalignant lung disease: COPD, only COPD; benign nodules, determined by biopsy; other, various types of lung diseases. C, receiver-operator characteristic curve for classification of samples shown in A and B. AUC, area under the curve. White circle, sensitivity-specificity value corresponding to classification score threshold of 0 (Showe et al 2009).

**Table 1**. Performance of 29 gene classifier on subclasses of patients and controls.

| # | *Subclass* | *Number of Samples* | *Accuracy by Class* |
|---|---|---|---|
| 1 | NSCLC | 137 | 91% |
| 2 | NHC | 91 | 80% |
| 3 | AC | 85 | 86% |
| 4 | LSCC | 42 | 98% |
| 5 | Nodules | 41 | 71% |
| 6 | COPD | 38 | 89% |
| 7 | Stage 1A | 48 | 83% |
| 8 | Stage 1B | 27 | 89% |
| 9 | Stage 1 | 75 | 85% |
| 10 | Stage 2 | 18 | 89% |
| 11 | Stage 3 | 39 | 100% |
| 12 | Stage 4 | 5 | 100% |

**Table 2**. Performance of 29 gene classifier based on smoking status.

| # | *Subclass* | *Number of Samples* | *Accuracy by Class* |
|---|---|---|---|
| 1 | Never Smokers | 24 | 88% |
| 2 | Former Smokers | 170 | 87% |
| 3 | Current Smokers | 34 | 79% |
| 4 | NSCLC: Never smokers | 9 | 89% |
| 5 | NSCLC: Former Smokers | 102 | 92% |
| 6 | NSCLC: Current Smokers | 26 | 85% |
| 7 | NHC: Never smokers | 15 | 87% |
| 8 | NHC: Former Smokers | 68 | 79% |
| 9 | NHC: Current Smokers | 8 | 63% |

**Table 3.** Positive predictive value and negative predictive value for 29-gene NSCLC classifier.

| Study | Sensitivity | Specificity | Prevalence | PPV | NPV |
|---|---|---|---|---|---|
| NSCLC vs NHC 29 gene classifier | 0.91 | 0.8 | 1% | 0.044 | 0.999 |
| | | | 5% | 0.193 | 0.994 |
| Spira et al., 2007 80 gene classifier | 0.8 | 0.84 | 1% | 0.048 | 0.998 |
| | | | 5% | 0.208 | 0.988 |
| LCGG Proposed biomarker | 0.8 | 0.7 | 1% | 0.026 | 0.997 |
| | | | 5% | 0.123 | 0.985 |

**Table 4.** 24-genes from classifier for NSCLC vs. non-malignant nodules

| # | Gene Description | Symbol | p-value | fold change |
|---|---|---|---|---|
| 1 | Peroxiredoxin 2 | PRDX2 | 0.00000 | 1.42 |
| 2 | RE1-silencing transcription factor | REST | 0.00034 | 1.40 |
| 3 | PREDICTED: similar to HLA class II histocompatibility antigen, DQ(W1.1) beta chain precursor (DQB1*0501), transcript variant 1 | LOC650557 | 0.00042 | -2.41 |
| 4 | Gamma-glutamyltransferase-like 3 (GGTL3), transcript variant 2 | GGTL3 | 0.00018 | 1.35 |
| 5 | Activating signal cointegrator 1 complex subunit 3 | ASCC3 | 0.00274 | 1.73 |
| 6 | Mediator complex subunit 16 | MED16 | 0.00027 | 1.22 |
| 7 | GTP-binding protein 10 (putative) | GTPBP10 | 0.00559 | -1.26 |
| 8 | Phosphatidic acid phosphatase type 2A | PPAP2A | 0.00355 | 1.20 |
| 9 | RAP1A, member of RAS oncogene family | RAP1A | 0.00018 | -1.35 |
| 10 | Brain and reproductive organ-expressed (TNFRSF1A modulator) | BRE | 0.00282 | 1.10 |
| 11 | PREDICTED: similar to Transcriptional regulator ATRX (X-linked helicase II) (X-linked nuclear protein) (XNP) | LOC652455 | 0.00029 | 1.47 |
| 12 | KIAA2026 | KIAA2026 | 0.00028 | -1.15 |
| 13 | Solute carrier family 11 (proton-coupled divalent metal ion transporters), member 1 | SLC11A1 | 0.00006 | -1.41 |
| 14 | Snurportin 1 | SNUPN | 0.00033 | -1.17 |
| 15 | Oxidative stress induced growth inhibitor family member 2 | OSGIN2 | 0.00204 | -1.09 |
| 16 | PREDICTED: spastic paraplegia 21 (autosomal recessive, Mast syndrome), transcript variant 3 (SPG21), mRNA | SPG21 | 0.02419 | 1.13 |
| 17 | Chromosome 19 open reading frame 12 | C19orf12 | 0.00382 | 1.43 |
| 18 | Ribosomal protein L6 | RPL6 | 0.00350 | -1.03 |
| 19 | Cas-Br-M (murine) ecotropic retroviral transforming sequence | CBL | 0.00001 | -1.18 |
| 20 | Rogdi homolog (Drosophila) | ROGDI | 0.00023 | -1.32 |
| 21 | 3'-phosphoadenosine 5'-phosphosulfate synthase 2 | PAPSS2 | 0.00341 | -1.34 |
| 22 | BAI1-associated protein 2 | BAIAP2 | 0.00046 | -1.38 |
| 23 | Family with sequence similarity 118, member A | FAM118A | 0.01598 | 1.94 |
| 24 | Triple functional domain (PTPRF interacting) | TRIO | 0.00001 | -1.31 |

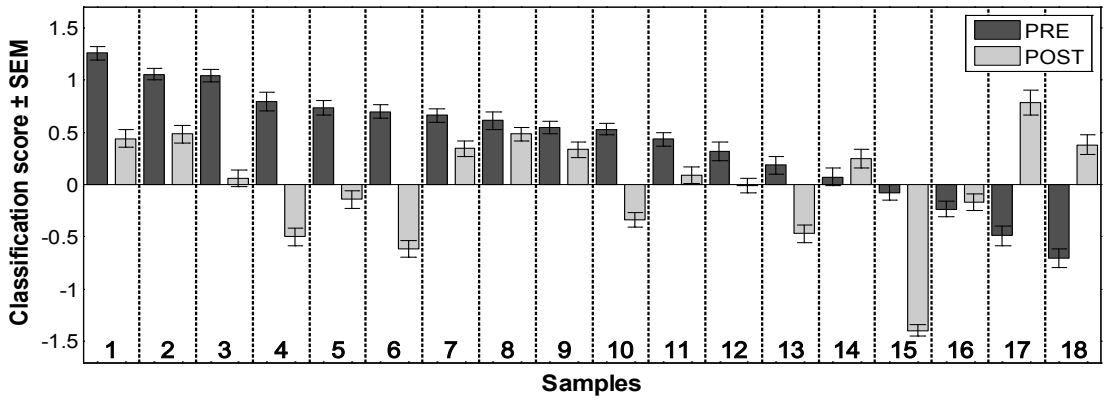


**Figure 2**: Classification Scores are altered by tumor removal. The samples are arranged as paired pre- and post-surgery samples to allow a comparison of the classification scores with the 29 gene diagnostic panel.

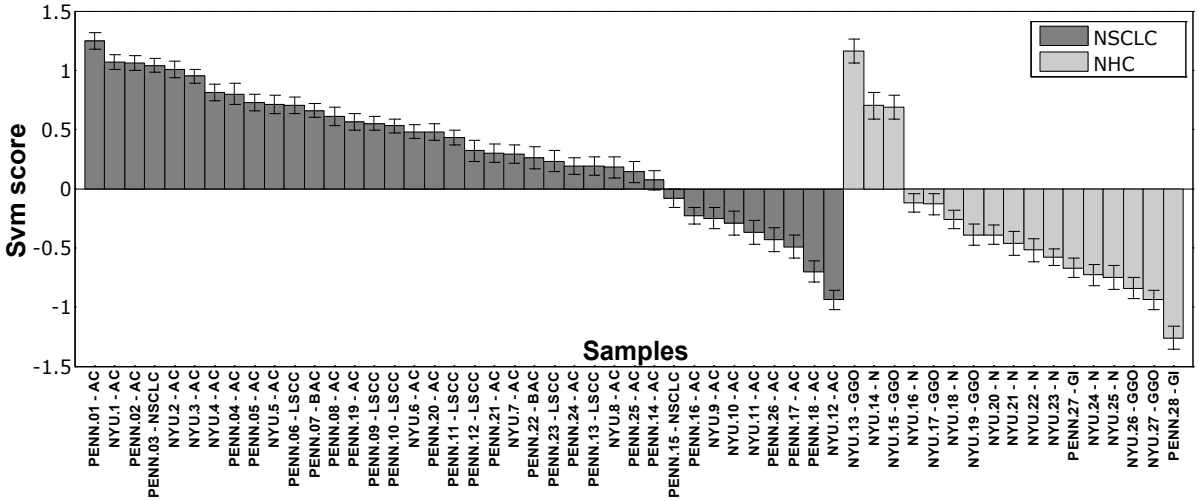


**Figure 3.** Application of the 29 gene NSCLC classifier to independent validation sets. PBMC derived RNA of lung cancer patients and controls collected at the New York University Lung Cancer Biomarker Center have labels prefaced by NYU. Lung cancer and control RNAs collected at Penn are prefaced by Penn. IDs that end in GGO= ground glass opacities, GI= granulomatous inflammation, AC=adenocarcinoma, LSCC=lung squamous cell carcinoma, NSCLC=non-small cell lung cancer, NHC=non-healthy control.

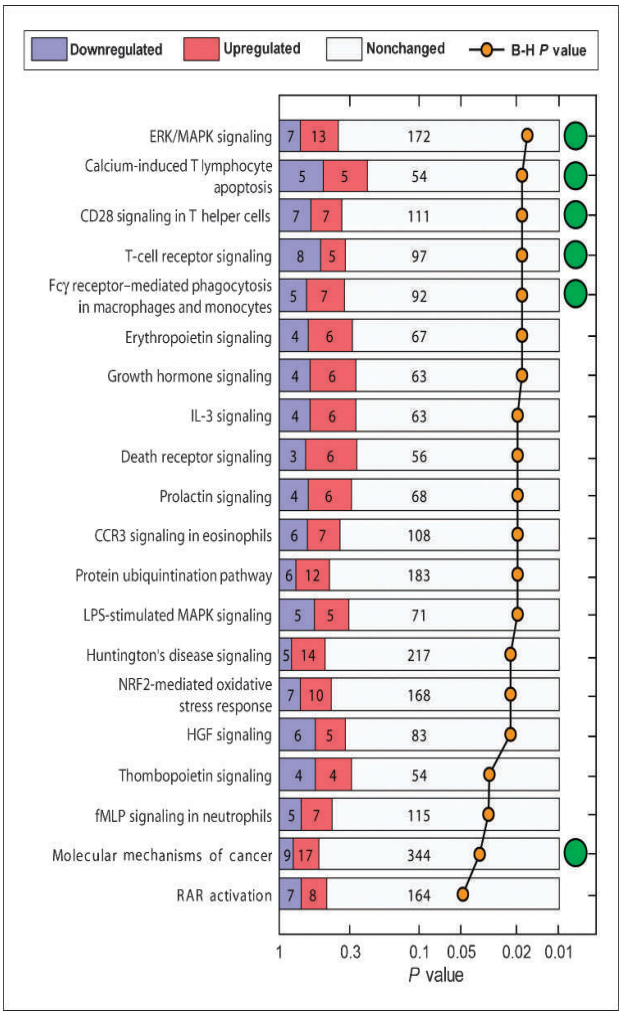


**Figure 4.** Significantly enriched canonical pathways from Ingenuity Pathway Analysis of the genes differentially regulated between NSCLC and NHC samples. Numbers in the bars, the number of genes in the pathway significantly higher in cancer(red) or lower in cancer (blue). B-H, Benjamini-Hochberg multiple testing correction. Green circles, pathways that were also enriched in NSCLC versus benign nodule comparison.

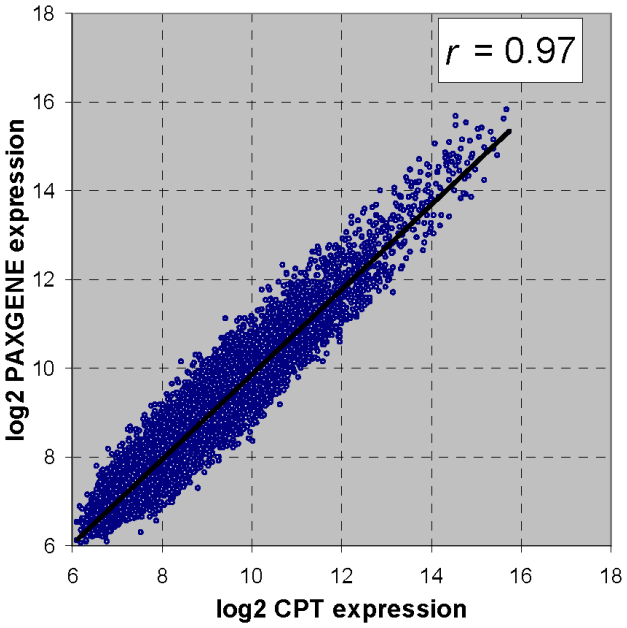


**Figure 5:** Comparison of gene expression in purified PBMC samples and whole blood RNA purified from PAXgene tubes showing a high degree of similarity between the global gene expression in samples collected using the 2 systems.

**18. Extent of Clinical Activities Initiated and Completed**. Items 18(A) and 18(B) should be completed for all research projects. If the project was restricted to secondary analysis of clinical data or data analysis of clinical research, then responses to 18(A) and 18(B) should be "No."

18(A) Did you initiate a study that involved the testing of treatment, prevention or diagnostic procedures on human subjects?

_____Yes
___X__ No

18(B) Did you complete a study that involved the testing of treatment, prevention or diagnostic procedures on human subjects?

_____Yes
___X__ No

**If "Yes" to either 18(A) or 18(B), items 18(C) – (F) must also be completed.** (Do NOT complete 18(C-F) if 18(A) and 18(B) are both "No.")

18(C) How many hospital and health care professionals were involved in the research project?

_____ Number of hospital and health care professionals involved in the research project

18(D) How many subjects were included in the study compared to targeted goals?

_____ Number of subjects originally targeted to be included in the study
_____ Number of subjects enrolled in the study

18(E) How many subjects were enrolled in the study by gender, ethnicity and race?

Gender:
_____Males
_____Females
_____Unknown

Ethnicity:
_____Latinos or Hispanics
_____Not Latinos or Hispanics
_____Unknown

Race:
_____American Indian or Alaska Native
_____Asian
_____Blacks or African American
_____Native Hawaiian or Other Pacific Islander
_____White
_____Other, specify:_____
_____Unknown

18(F) Where was the research study conducted? (List the county where the research study was conducted. If the treatment, prevention and diagnostic tests were offered in more than one county, list all of the counties where the research study was conducted.)

**19. Human Embryonic Stem Cell Research.** Item 19(A) should be completed for all research projects. If the research project involved human embryonic stem cells, items 19(B) and 19(C) must also be completed.

19(A) Did this project involve, in any capacity, human embryonic stem cells?
_____Yes
__X___ No

19(B) Were these stem cell lines NIH-approved lines that were derived outside of Pennsylvania?

_____Yes
_____ No

19(C) Please describe how this project involved human embryonic stem cells:

**20. Articles Submitted to Peer-Reviewed Publications**.

20(A) Identify all publications that resulted from the research performed during the funding period and that have been submitted to peer-reviewed publications. Do not list journal abstracts or presentations at professional meetings; abstract and meeting presentations should be listed at the end of item 17. **Include only those publications that acknowledge the Pennsylvania Department of Health as a funding source** (as required in the grant agreement). List the title of the journal article, the authors, the name of the peer-reviewed publication, the month and year when it was submitted, and the status of publication (submitted for publication, accepted for publication or published.). Submit an electronic copy of each publication, listed in the table, in a PDF version 5.0.5 format, 1,200 dpi. Filenames for each publication should include the number of the research project, the last name of the PI, the number of the publication and an abbreviated research project title. For example, if you submit two publications for PI Smith for the "Cognition and MRI in Older Adults" research project (Project 1), and two publications for PI Zhang for the "Lung Cancer" research project (Project 3), the filenames should be:
Project 1 – Smith – Publication 1 – Cognition and MRI
Project 1 – Smith – Publication 2 – Cognition and MRI
Project 3 – Zhang – Publication 1 – Lung Cancer
Project 3 – Zhang – Publication 2 – Lung Cancer
If the publication is not available electronically, provide 5 paper copies of the publication.

**Note:** The grant agreement requires that recipients acknowledge the Pennsylvania Department of Health funding in all publications. Please ensure that all publications listed acknowledge the Department of Health funding. If a publication does not acknowledge the funding from the Commonwealth, do not list the publication.

| Title of Journal Article: | Authors | Name of Peer-reviewed Publication: | Month and Year Submitted: | Publication Status (check appropriate box below): |
|---|---|---|---|---|
| 1. Gene Expression Profiles in Peripheral Blood Mononuclear Cells Can Distinguish Patients with Non-Small Cell Lung Cancer from Patients with Nonmalignant Lung Disease | **Showe, Michael K.,** Vachani, Anil, **Kossenkov, Andrew V.,** Yousef, Malik, **Nichols, Calen**, Nikonova, Elena V., **Chang, Celia**, Kucharczuk, John, Tran, Bao, | Cancer Research | | ☐Submitted<br>☐Accepted<br>☒ Published |

| | | | | |
|---|---|---|---|---|
| | Wakeam, Elliot, Yie, Ting An, Speicher, David, Rom, William N., Albelda, Steven, **Showe, Louise C**: | | | |

20(B) Based on this project, are you planning to submit articles to peer-reviewed publications in the future?

Yes ____X_____ No_____

If yes, please describe your plans:

We are in the process of submitting the 2 manuscripts listed below:

1.Differential Expression Of Coding And Non-Coding Genes in PBMCs Of NSCLC Patients After Tumor Removal. Andrew V. Kossenkov[1], Anil Vachani[2], Celia Chang[1], Calen Nichols[1], Shere Billouin[1], WenHwai Horng[1], William Rom, Steven Albelda[2], Michael K. Showe[1], and Louise C. Showe[1]

2. Predicting Outcome in NSCLC Patients Based on Gene Expression in Peripheral Blood. Anil Vachani, Andrew V. Kossenkov, William Rom, Steven Albelda[2], Michael K. Showe[1], and Louise C. Showe[1]

**21. Changes in Outcome, Impact and Effectiveness Attributable to the Research Project.** Describe the outcome, impact, and effectiveness of the research project by summarizing its impact on the incidence of disease, death from disease, stage of disease at time of diagnosis, or other relevant measures of outcome, impact or effectiveness of the research project. If there were no changes, insert "None"; do not use "Not applicable." Responses must be single-spaced below, and no smaller than 12-point type. DO NOT DELETE THESE INSTRUCTIONS. There is no limit to the length of your response.

The use of peripheral blood to diagnose early stage NSCLC disease promises to be a major step forward in improving lung cancer survival by detecting malignant tumors in high-risk patients (i.e. smokers and ex-smokers) at an early stage, while they are still surgically resectable.

**22. Major Discoveries, New Drugs, and New Approaches for Prevention Diagnosis and Treatment**. Describe major discoveries, new drugs, and new approaches for prevention, diagnosis and treatment that are attributable to the completed research project. If there were no major discoveries, drugs or approaches, insert "None"; do not use "Not applicable." Responses must be single-spaced below, and no smaller than 12-point type. DO NOT DELETE THESE INSTRUCTIONS. There is no limit to the length of your response.

We demonstrated that (i) the gene expression in peripheral blood could be used to diagnose NSCLC with accuracies similar to more invasive procedures like bronchoscopy or methods which assay gene expression in sputum; (ii) that these blood samples contain gene expression data that is predictive of outcome; and (iii) that analysis of pre and post surgery blood samples from patients with NSCLC may contain information that could be predictive of outcome and recurrence.

**23. Inventions, Patents and Commercial Development Opportunities.**

23(A) Were any inventions, which may be patentable or otherwise protectable under Title 35 of the United States Code, conceived or first actually reduced to practice in the performance of work under this health research grant? Yes__x_____ No_____

If "Yes" to 23(A), complete items a – g below for each invention. (Do NOT complete items a - g if 23(A) is "No.")

a. Title of Invention: Method for Diagnosing Lung Cancers using Gene Expression Profiles in Peripheral Blood Mononuclear Cells

b. Name of Inventor(s): Louise C. Showe, Michael K. Showe[1], Andrew V. Kossenkov, Anil Vachani, Steven Albelda

c. Technical Description of Invention (describe nature, purpose, operation and physical, chemical, biological or electrical characteristics of the invention):

Patent PCT/US2008/013450 ("Method for Diagnosing Lung Cancers using Gene Expression Profiles in Peripheral Blood Mononuclear Cells") was filed 12/05/2008. National phase applications have been filed in the United States and Europe. This patent application, jointly owned by The Wistar Institute and the University of Pennsylvania, claims compositions for and methods of diagnosing and staging lung cancer using purified white blood cells (PBMC) and includes multiple claims of sets of genes used in the diagnosis and/or staging of the cancer. The European patent application was filed at the request of a local venture firm that is interested in commercializing this test.

d. Was a patent filed for the invention conceived or first actually reduced to practice in the performance of work under this health research grant?
Yes__x__ No___

If yes, indicate date patent was filed:

e. Was a patent issued for the invention conceived or first actually reduced to practice in the performance of work under this health research grant?
Yes X No___

If yes, indicate number of patent, title and date issued:
Patent number: Patent PCT/US2008/013450

Title of patent: "Method for Diagnosing Lung Cancers using Gene Expression Profiles in Peripheral Blood Mononuclear Cells
Date issued: 07/14/2010

f. Were any licenses granted for the patent obtained as a result of work performed under this health research grant? Yes ___ No _X___

If yes, how many licenses were granted?_____

g. Were any commercial development activities taken to develop the invention into a commercial product or service for manufacture or sale? Yes___ No____X

If yes, describe the commercial development activities:

23(B) Based on the results of this project, are you planning to file for any licenses or patents, or undertake any commercial development opportunities in the future?

Yes ___X_____ No_____

If yes, please describe your plans:

We have ongoing licensing agreements with Phoenix Ltd.

**24. Key Investigator Qualifications.** Briefly describe the education, research interests and experience and professional commitments of the Principal Investigator and all other key investigators. In place of narrative you may insert the NIH biosketch form here; however, please limit each biosketch to 1-2 pages. *For Nonformula grants only – include information for only those key investigators whose biosketches were not included in the original grant application.*

## BIOGRAPHICAL SKETCH

Provide the following information for the key personnel and other significant contributors in the order listed on Form Page 2.
Follow this format for each person. **DO NOT EXCEED FOUR PAGES.**

| NAME | POSITION TITLE |
|---|---|
| **SHOWE, Louise C.** | **Professor** |
| eRA Commons Username<br>**LSHOWE** | |

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, and include postdoctoral training.)*

| INSTITUTION AND LOCATION | DEGREE *(if applicable)* | YEAR(s) | FIELD OF STUDY |
|---|---|---|---|
| Wilkes College, Wilkes Barre, PA | B.A. | 1958-1962 | Biology |
| Temple University, Philadelphia, PA | M.A. | 1968-1969 | Developmental Biology |
| University of Pennsylvania, Philadelphia, PA | Ph.D. | 1973-1975 | Biology |

**A. Personal Statement**:

Dr. Showe is a professor and member of both the Molecular Oncogenesis and Immunology programs at Wistar indicating her interest in the interactions of cancer and the immune system and the application of genomic, bioinformatics' and proteomic approaches to solving complex problems. She established the first functioning Genomics Core in the Philadelphia area and continues as the Scientific Director. She is well respected in the Genomics field and was one of the first 10 recipients of the NCI Director's Challenge grants to develop molecular diagnostics for cancer using microarray studies. During this period she also developed a strong bioinformatics group within her lab working on applying and developing new approaches to high dimensional data analysis. In 2008 Wistar established a Center for Systems and Computational Biology of which she is an Associate Director of this center. Recent work has focused on the detection of a cancer signature in the peripheral immune system of patients with solid tumors, work that directly relates to the proposed studies. This work focuses on non-small cell lung cancer (NSCLC) and a recent paper in Cancer Research (Showe et al 2009) describes the development of a 29 gene signature identified in peripheral blood that distinguishes patients with early stage NSCLC from patients with smoking related non-malignant lung disease with 86% accuracy using a Support Vector Machine (SVM)( ). They also demonstrated that the SVM derived cancer signature is diminished in post surgery in 76% of a small subset of 18 pre vs. post surgery PBMC samples. In a paper soon to be submitted the differences in the PBMC signatures in the pre vs post surgery samples are further explored at both the gene and miRNA levels and the correlation of gene and of target genes and their miRNAs are explored. The use of matched pre-post samples provides an opportunity to address patient specific as well as shared characteristics of the effects of the tumor presence on the immune system**.**

**B. Positions and Honors:**

Research Experience:

| | |
|---|---|
| 1975-1978 | Postdoctoral Studies, Biozentrum der Universitat, Basel, Switzerland |
| 1975-1978 | Post-doctoral Fellowships: NATO. NIH, Swiss Science Foundatgion |
| 1979-1980 | Research Associate, Department of Biochemistry & Biophysics, University of Pennsylvania School of Medicine |
| 1980-1982 | Research Assistant Professor, Department of Biochemistry & Biophysics, University of Pennsylvania School of Medicine |
| 1982-1983 | Research Assistant Professor, Department of Pediatrics, CHOP, University of Pennsylvania |
| 1983-1989 | Assistant Professor, The Wistar Institute, Philadelphia, PA |
| 1990-2006 | Associate Professor, The Wistar Institute, Philadelphia, PA |
| 1999-present | Scientific Director, Genomics Facility, The Wistar Institute Cancer Center, Philadelphia, PA |
| 2003-2007 | Wistar Associate Professor of Dermatology, University of Pennsylvania |
| 2007-present | Professor, Molecular Oncogenesis and Immunology, The Wistar Institute, Philadelphia, PA |
| 2007-present | Wistar Professor of Dermatology, University of Pennsylvania |
| 2008-present | Associate Director Genomics, Center for Computation and Systems Biology, The Wistar Institute |
| 2009-present | Board of Directors PA Cancer Control Consortium |

Advisory Activities:

| | |
|---|---|
| 1999- | NCI Clinical Studies, Biomarkers, Cancer Prognosis and Prediction, IMAT etc. |
| 2001- | Reviewer Cancer Research, Blood, Hematologica, JID, NAR, BMC, Clinical Cancer Research, Bioinformatics |
| 2001-2007 | Wistar Training Committee, Editorial Committee, |
| 2004- | Wistar Personnel Committee-Chair |
| 2002-2008 | PA Cancer Control Consortium, PAC3-Research Co-Chair |
| 2002- | Greater Philadelphia Bioinformatics Consortium-Executive Committee |

| | |
|---|---|
| 2002- | Philadelphia Summer Institute for Bioinformatics, Board of Advisors |
| 2003-2006 | Cancer Biomarkers Study Section |
| 2005-2008 | Penn Genomics Institute Oversight Committee |
| 2006- | Wistar Internal Steering Committee, Staff Council, Faculty Recruitment |

**B. Selected Peer-Reviewed Publications (in chronological order):**

1. **Showe, L.,** Fox, F.E., Williams, D., Au, K., Niu, Z., and Rook, A.H. 1999. Depressed interleukin-12 (IL-12) mediated signal transduction in T-cells from patients with Sezary Syndrome is associated with the absence of IL-12 Receptor □2 mRNA and reduced levels of STAT-4. J. Immunol. 163:4073-79.
2. Heath, V.L., **Showe, L.**, Crain, C., Barrat, F.J., Trinchieri, G., O'Garra, A. 2000. Cutting Edge: Ectopic expression of the IL-12 receptor-β2 in developing and committed Th2 cells does not affect the production of IL-4 or induce the production of IFN-γ. J. Immunol. 164:2861-2865.
3. Kari, L., Loboda, A., Nebozhyn, M., Rook, A.H., Vonderheid, E.C., Nichols, C., Virok, D., Chang, C., Horng, W-H., Johnston, J., Wysocka, M., Showe, M.K., and **Showe, L.C**. 2003. Classification and prediction of survival in patients with the leukemic phase of cutaneous T-cell lymphoma. J. Exp. Med. 197:1477-1488.
4. Virok, P., Loboda, A., Kari, L., Nebozhyn, M., Chang, C., Berencsi, K., Showe, M.K., Gonczol, E., and **Showe, L**. 2003. Infection of U937 monocytic cells with *Chlamydia pneumoniae* induces extensive changes in host cell gene expression. J. Infect. Dis. 188:1310-1321.
5. Nebozhyn, M., Loboda, A., Kari, L., Rook, A.H., Vonderheid, E.C., Lessin, S., Berger, C., Edelson, R., Nichols, C., Yousef, M., Gudipati, L., Shang, M., Showe, M.K. and **Showe, L.C**. 2006. Quantitative PCR on 5 genes reliably identifies CTCL patients with 5-99% circulating tumor cells with 90% accuracy. Blood. 15: 107: 3189-3196.
6. Yousef, M., Nebozhyn, M., Shatkay, H., Kanterakis, S., **Showe, L.C.**, and Showe, M.K. 2006. Combining multi-species genomic data for microRNA identification using a Naïve Bayes classifier machine learning for identification of microRNA genes. Bioinformatics 22: 1325-1334.
7. Anil Vachani, Michael Nebozhyn, Sunil Singhal, Linda Alila, Elliot Wakeam, Ruth Muschel, Charles A. Powell, Patrick Gaffney, Bhuvanesh Singh, Marcia S. Brose, Leslie A. Litzky, John Kucharczuk, Larry R. Kaiser, J. Stephen Marron, Michael K. Showe, Steven M. Albelda, **Louise C. Showe**. 2007. A 10-Gene Classifier for Distinguishing Head and Neck Squamous Cell Carcinoma and Lung Squamous Cell Carcinoma. Clin. Cancer Res. 13: 2905-2915.
8. Malik Yousef, Segun Jung**, Louise C. Showe** and Michael K. Showe. 2007. Recursive Cluster Elimination (RCE) for Classification and Feature Selection from Gene Expression Data. BMC Bioinformatics 8:144.
9. Malik Yousef, Segun Jung, Andrew V. Kossenkov, **Louise C. Showe**, and Michael K. Showe. 2007. Naïve Bayes for microRNA target predictions—machine learning for microRNA targets. Bioinformatics 23: 2987-2992.
10. Malavika S Giri, Michael Nebozyhn, Shenoa Creer, Calen Nichols, Andrea Raymond, Aidan Hancock, Shannon Southhall, Andrea S Foulkes, Cele Gallo, Karam Mounzer, Guido Silvestri, **Louise C Showe**, Luis J Montaner, Malik Yousef, Jay Kostman, Ronald. G. Collman. 2009. Circulating monocytes in HIV-1-infected viremic subjects exhibit an antiapoptosis gene signature and virus- and host-mediated apoptosis resistance. Journal of Immunology 182:4459-70.
11. Malik Yousef, Mohamed Ketany, Larry Manevitz, **Louise C Showe** and Michael K Showe. Classification and biomarker identification using gene network modules and support vector machines. BMC Bioinformatics 2009, 10:337
12. Kiranmai Gumireddy, Anping Li, Phyllis A. Gimotty, Andres J. Klein-Szanto, **Louise C. Showe**, Dionyssios Katsaros, George Coukos, Lin Zhang & Qihong Huang. KLF17 is a negative regulator of epithelial–mesenchymal transition and metastasis in breast cancer. Nat Cell Biol. 2009 Nov;11(11):1297-304. Epub 2009 Oct 4.
13. Blankenhorn EP, Bryan G, Kossenkov AV, Clark LD, Zhang XM, Chang C, Horng W, Pletscher LS, Cheverud JM, **Showe LC**, Heber-Katz E. Genetic loci that regulate healing and regeneration in LG/J and SM/J mice. Mamm Genome. 2009 Sep 16. [Epub ahead of print]
14. Qin H, Chan MW, Liyanarachchi S, Balch C, Potter D, Souriraj IJ, Cheng AS, Agosto-Perez FJ, Nikonova EV, Yan PS, Lin HJ, Nephew KP, Saltz JH, **Showe LC**, Huang TH, Davuluri RV. An integrative ChIP-chip and gene expression profiling to model SMAD regulatory modules. BMC Syst Biol. 2009 Jul 17;3:73
15. Showe, Michael K., Vachani, Anil, Kossenkov, Andrew V., Yousef, Malik, Nichols, Calen, Nikonova, Elena V., Chang, Celia, Kucharczuk, John, Tran, Bao, Wakeam, Elliot, Yie, Ting An, Speicher, David, Rom, William N., Albelda, Steven, **Showe, Louise C.** Gene Expression Profiles in Peripheral Blood Mononuclear Cells Can Distinguish Patients with Non-Small Cell Lung Cancer from Patients with Nonmalignant Lung Disease. Cancer Res 2009 69: 9202-9210