AWARD NUMBER: W81XWH-14-1-0263

TITLE: Early Detection of NSCLC Using Stromal Markers in Peripheral Blood

PRINCIPAL INVESTIGATOR: Dingcheng Gao

CONTRACTING ORGANIZATION: Weill Cornell Medical College
New York, NY 10065

REPORT DATE: September 2016

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE
September 2016

2. REPORT TYPE
Annual

3. DATES COVERED
09/01/2015 – 08/31/2016

4. TITLE AND SUBTITLE

Early Detection of NSCLC Using Stromal Markers in Peripheral Blood

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-14-1-0263

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Dingcheng Gao

E-Mail: dig2009@med.cornell.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Weill Cornell Medical College 1300 York Ave, New York, Ny 10065

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
A recent screening trial showed that the use of low dose computed tomography (LDCT) resulted in a 20% reduction in lung cancer mortality, however there was a 96% false positive rate associated with LDCT. Thus, there is an immediate clinical need to develop a diagnostic biomarker that would select patients with CT detected nodules for further testing. The ease with which blood can be sampled makes it a logical choice in which to discover diagnostic biomarkers, however the clinical utility of tumor derived proteins, miRNA or circulating tumor cells as blood-based biomarkers has been limited. In this proposal, instead of tumor-derived biomarkers, we will focus on host response to tumor growth. It has been well documented that tumor growth systemically stimulates and mobilizes BM-derived hematopoietic cells to the tumor bed to establish a permissive microenvironment. Preliminary studies in our lab have shown that in lung cancer patients, the circulating myeloid cells are transcriptionally altered and the alteration is tumor dependent. The specific transcriptomic signature of circulating myeloid cells may provide us unique resources for lung cancer biomarker discovery. Therefore, we hypothesized that the circulating BM-derived myeloid cells carry specific transcriptomic signature, which may be useful for early lung cancer diagnosis. The specific aims are: **Aim 1.** To identify a NSCLC-dependent transcriptomic signature in circulating myeloid cells. **Aim 2.** To validate the diagnostic value of the specific gene signatures of circulating myeloid cells in NSCLC patients with lung nodules.

15. SUBJECT TERMS

Nothing listed

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON USAMRMC
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(include area code)*
Unclassified	Unclassified	Unclassified	Unclassified	10	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Table of Contents

Page

1. Introduction……………………………………………………….1

2. Keywords……………………………………………………….1

3. Accomplishments………..…………………………………………...1

4. Impact…………………………...…………………………………5

5. Changes/Problems...….………………………………………………5

6. Products…………………………………….……….….……………6

7. Participants & Other Collaborating Organizations……………6

8. Special Reporting Requirements……………………………………7

9. Appendices……………………………………………………………7

1. **INTRODUCTION:**

A recent screening trial showed that the use of low dose computed tomography (LDCT) resulted in a 20% reduction in lung cancer mortality, however there was a 96% false positive rate associated with LDCT. Thus, there is an immediate clinical need to develop a diagnostic biomarker that would select patients with CT detected nodules for further testing. The ease with which blood can be sampled makes it a logical choice in which to discover diagnostic biomarkers, however the clinical utility of tumor derived proteins, miRNA or circulating tumor cells as blood-based biomarkers has been limited. In this proposal, instead of tumor-derived biomarkers, we will focus on host response to tumor growth. It has been well documented that tumor growth systemically stimulates and mobilizes BM-derived hematopoietic cells to the tumor bed to establish a permissive microenvironment. Preliminary studies in our lab have shown that in lung cancer patients, the circulating myeloid cells are transcriptionally altered and the alteration is tumor dependent. The specific transcriptomic signature of circulating myeloid cells may provide us unique resources for lung cancer biomarker discovery. Therefore, we proposed to identify a NSCLC-dependent transcriptomic signature in circulating myeloid cells and then validate the diagnostic value of the specific gene signatures of circulating myeloid cells in NSCLC patients. The proposed study, if succeed, will provide novel strategies and approaches for early detection of lung cancer.

2. **KEYWORDS:**

None small cell lung cancer (NSCLC), biomarker, circulating myeloid cells, flow cytometry, RNA-sequencing, expression profiling.

3. **ACCOMPLISHMENTS:**

- **What were the major goals of the project?**

Specific Aim 1: To identify a NSCLC-dependent transcriptomic signature in circulating myeloid cells. (Proposed to be accomplished during the first year)

Major Task 1. Lung cancer signature gene optimization

Subtask 1: Patient recruitment including pre- and post- surgery patients, and COPD patients

Subtask 2: Flow cytometry sorting of circulating myeloid cells.

Subtask 3: RNA-Sequencing

Subtask 4: RNA-seq data analysis

Subtask 5: Feasible RT-PCR array assay development

Specific Aim 2: To validate the diagnostic value of the specific gene signatures of circulating myeloid cells in patients with lung nodules. (Proposed to be accomplished during the second year)

Major Task 2: Lung cancer signature diagnostic value validation

Subtask 1: Recruitment of patients with positive lung nodules by CT-Scan

Subtask 2: Flow cytometry sorting of circulating myeloid cells

Subtask 3: RT-PCR array and data analysis with clinical outcomes

- **What was accomplished under these goals?**

For this reporting period, we followed the proposal, accomplished the patient recruitment, flow cytometry sorting of circulating myeloid cells, RNA-sequencing of the samples. During the RNA-seq data analysis, we encountered challenges of batch differences, patient gender differences, great variations in identified genes expressions even within sorted subpopulations. We are currently working closely with our bioinformatics collaborators. As reliable biomarkers for lung cancer are identified, we will move forward to re-value it in patient samples, which have been bio-banked in our lab. We have required a No-Cost-Extension period of one year and expect to be able to overcome these difficulties in the extension period.

Major Task 1. Lung cancer signature gene optimization

Subtask 1: Patient recruitment including pre- and post- surgery patients, and COPD patients

During this reporting period, we have recruited 23 NSCLC patients and collected their peripheral blood before and after the surgical removal of the primary lung tumor. Peripheral blood was also collected from 6 patients with benign nodules to serve as non-tumor control group. All blood samples were sorted via flow cytometry into IMMCs and polymorphonuclear neutrophils. An unfractionated whole white blood cells aliquot was also retained. Total RNA was extracted and RNA sequencing (poly-A selected, single-read, 51 bp, 6 samples per lane) was performed using an Illumina sequencer.

Subtask 2: Flow cytometry sorting of circulating myeloid cells.

With the peripheral blood, we have performed cytometry to isolate CD11b+CD33- neutrophils and CD11b+CD33+ monocytic myeloid cells. The sorting strategies of these myeloid cells have been well-established (Fig. 1, a representative patient sample from the previous report). Consistent as we found in the preliminary study, the absolute number and the percentage of myeloid subpopulations showed broad variations between patients. After sorting of subpopulation, such variations in cell numbers was supposed to be normalized per cell and further justified the usage of flow cytometry sorting to isolate myeloid cells for the gene expression profiling analysis.

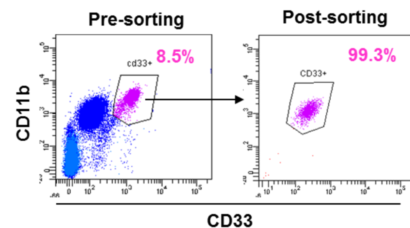

Figure 1. The purity of CD11b+CD33+ myeloid cells in the pre- and post-sorting samples.

Subtask 3: RNA-Sequencing

We extracted RNA from sorted cells using the mirVana kit (Life Technologies). Using the TruSeq RNA sample preparation kit (Illumina, Inc) cDNA libraries was constructed. We performed 51bp single read with HiSeq machines in the Genome Sequencing Facility at WCMC. Short reads (after FastQC quality control) were mapped to hg19 using TopHat and expression levels quantified using CuffLinks. Gene expression level (FPKM) was determined using DEseq and LIMMA. We applied

the RNA-seq by single-read for 51 cycles and pooled 6 samples per lane. This strategy has given us reliable sequence-reading with deep enough coverage of the transcriptome.

Subtask 4: RNA-seq data analysis.

With the RNA-sequencing results from 23 paired pre- and post-surgical samples, we have first performed clustering analysis. Consistent with our preliminary results, Both CD33+ monocytic myeloid cells and CD33- neutrophils showed a unique gene expression profile which was distinguishable from that of total cells (Fig. 2). These results indicated the high reliability and reproducibility of expression profiles from different myeloid subpopulation cells.

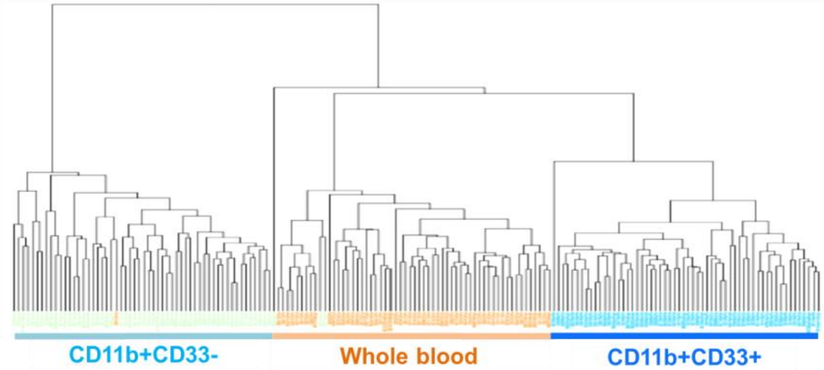

Figure 2. Clustering analysis of RNA-sequencing results of the sorted monocytes (CD11b+CD33+), neutrophils (CD11b+CD33-) and the unsorted whole blood. Differential gene expression profiles were detected from different subtypes of myeloid cells which were distinguishable from the unsorted whole blood samples.

To identify the candidate genes that may serve as biomarkers for lung cancer, we aimed to compare pre versus post-surgery samples. Very few genes that are up- or down- regulated by >1.2 fold with adjusted p value <0.05 and FPKM value ≥5 were identified. There is no differential expressed gene identified with comparison of whole blood samples. With the sorted neutrophils (CD11b+CD33-) samples, 4 genes (GPI, ABCC1, ESYT1 and TAF15) were identified. With the sorted monocyte samples, 1 gene (KLRF1) were identified. Of note, this is in contrast with the 203 genes in neutrophils and 22 genes in monocytes that we identified using 15 paired pre versus post-surgical samples. In addition, these five genes were not included in the original list, which make us doubt about the reliability of these results.

Further analysis of KLRF1 expression in sorted monocytes, we found that its expression was detected as 183.9±126.4 and 149.5±108.3 in pre- and post-surgical samples, respectively. There is big variation between patients (Fig. 3), though the trend of down-regulation after removal of lung tumor was also detected. Analyses of other candidate genes identified in neutrophils showed similar results.

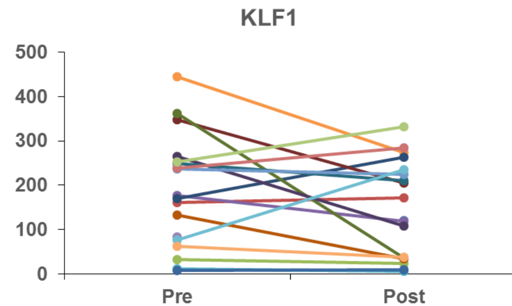

Figure 3. The expression of KLF1 in sorted monocytes from pre- and post-surgical blood samples of lung cancer patients.

To further clarify what cause the dramatic loss of candidate genes, we performed two-way unsupervised clustering analysis with all samples. Consistent with previous one-way clustering analysis, different cell types (neutrophils, monocytes and unsorted whole blood cells) showed clear clustering in the two-way analysis (Fig. 4A). However, as we analyzing samples from a single cell type, a batch difference was clearly detected from the samples that we submitted for sequencing at different time (Fig. 4B). This will cause the big variation in the sequencing readouts, that led to the dramatic loss of candidate genes when will

pool all samples together for the final analysis. This also ask us to improve our bioinformatics analysis strategy to correct this systematic error.

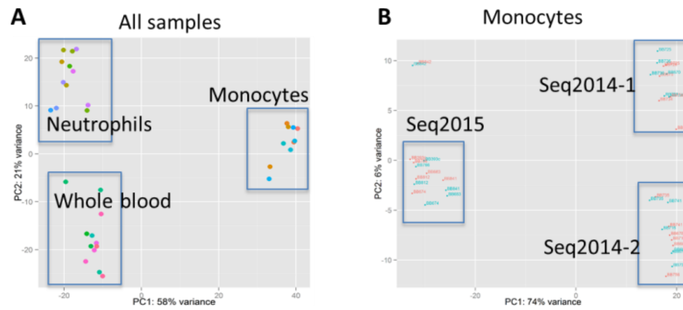

Figure 4. Two-way unsupervised clustering analysis of RNA-seq results. A, Plot of all samples analysis showing the clustering of different cell types. B, Plot of monocyte sample analysis showing the clustering of samples from different batches.

Major Task 2: Lung cancer signature diagnostic value validation

Subtask 1: Recruitment of patients with positive lung nodules by CT-Scan

To facilitate further confirmation of the identified candidate genes as biomarker for lung cancer detection, we have biobanked blood samples from 120 patients in total including lung cancer patients and patients with positive lung nodules by CT-Scan. Total RNA has been extracted from the samples and preserved in biobank for further analysis.

Subtask 2: Flow cytometry sorting of circulating myeloid cells

Following the same strategy (Fig. 1), we have sorted CD11b+CD33- neutrophils and CD11b+CD33+ monocytic myeloid cells from blood samples of 6 patients with confirmed benign lung nodules. These samples will serve as non-tumor controls in RT-PCR analysis proposal in subtask 3.

- **What opportunities for training and professional development has the project provided?**

  Nothing to Report.

- **How were the results disseminated to communities of interest?**

  Nothing to Report.

- **What do you plan to do during the next reporting period to accomplish the goals?**

In the next report period, we will first focus on the bioinformatic analysis of the RNA-seq data. While we confirm the identified genes (5 genes in total) are related to lung cancer. We will design an RT-PCR array to confirm the diagnostic value of these candidate genes with biobanked blood samples.

Given the high variation of candidate gene expression in different cancer patients, we will improve our bioinformatics analysis strategy. We plan to introduce cell type markers (CD45, CD11b and CD33) together with house-keeping gene markers (GAPDH, 18SrRNA, ACTB, and B2M) to normalize candidate gene expression in total whole blood RNA samples. This strategy will allow us correct the system error from total cell numbers and percentages of subtype population cells. The modified following tasks will be:

Major Task 2: Lung cancer signature diagnostic value validation

Subtask 3: Feasible RT-PCR array assay development.

Subtask 4: RT-PCR array and data analysis with clinical outcomes.

4. **IMPACT:**

- **What was the impact on the development of the principal discipline(s) of the project?**

The persistent poor survival of lung cancer patients is largely attributable to the late stage at diagnosis. New biomarkers for early detection are urgently required in the clinic. However, discovery of biomarkers using peripheral blood is challenging because tumor-specific markers are usually expressed in low concentrations, diluted in a milieu of other abundant proteins and likely to be missed. To overcome this hurdle, instead of focusing on tumor-derived biomarkers, we will analyze the host responses to the tumor growth. The abundance of circulating myeloid cells, which we know play important roles in tumor growth, may provide a unique source for novel NSCLC biomarker discovery.

During the first year of the project, we have recruited enough NSCLC patients as proposed for pre- and post-surgery comparison analysis. We have optimized sorting strategies for circulating myeloid cells, which may possess unique expression signature for early lung cancer detection.

During the second year of the project, we have accomplished the patient recruitment, flow cytometry sorting of circulating myeloid cells, RNA-sequencing of the samples. RNA-seq data analysis was performed with challenges. We are currently working closely with our bioinformatics collaborators to solve the problems and expect to fulfill the proposal with the No-Cost-Extension period.

- **What was the impact on other disciplines?**

  Nothing to Report.

- **What was the impact on technology transfer?**

  Nothing to Report.

- **What was the impact on society beyond science and technology?**

  Nothing to Report.

5. **CHANGES/PROBLEMS:**

We have encountered challenges in bioinformatics analysis of RNA-seq data. Very limited number of genes (4 from sorted neutrophils, 1 from sorted monocytes) were identified. Also, high variations in candidate genes expression were detected between patients. This might be due to the batch difference in RNA-seq results. Extra experiments will be needed to confirm the candidate genes as biomarker for lung cancer diagnosis. We plan to improve our RT-PCR array assay by including cell type markers together with house-keeping gene markers to normalize the variabilities in total cell numbers and percentages in cell types from different patients. This will also allow us to to perform the assay with the whole blood samples as well as with the sorted

subpopulations of blood samples. We have inquired a Non-Cost-Extension period (1 year) to finish this proposal.

6. **PRODUCTS:**

Nothing to Report.

7. **PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS**

- **What individuals have worked on the project?**

Name:	*Dingcheng Gao*
Project Role:	*PI*
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	*12*
Contribution to Project:	*Dr. Gao has overseen the ongoing project, performed work in lung cancer biomarker discovery by combining flow cytometry and RNA-sequencing techniques.*
Funding Support:	

Name:	*Nasser Altorki*
Project Role:	*Co-PI*
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	12
Contribution to Project:	*Dr. Altorki has guided the collection of patient samples, cooperate with pathologist and lab members for biobanking management with 5% efforts.*
Funding Support:	

Name:	*Oliver Elemento*

Project Role:	*Co-PI*
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	12
Contribution to Project:	*Dr. Elemento has been in charge of bioinformatics analysis of the project with 3.8% efforts.*
Funding Support:	

Name:	*Cathy Spinelli*
Project Role:	*Clinical coordinator*
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	12
Contribution to Project:	*Ms. Spinelli has supported the collection and biobanking of patient blood samples with 5% efforts.*
Funding Support:	

Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?

Nothing to Report

What other organizations were involved as partners?

Nothing to Report