AWARD NUMBER: W81XWH-14-1-0263

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

PRINCIPAL INVESTIGATOR: Dingcheng Gao

CONTRACTING ORGANIZATION: Joan & Sanford I. Weill Medical College of Cornell University
New York, NY 10065

REPORT DATE: September 2015

TYPE OF REPORT: Annual

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

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
September 2015	Annual Progress Report	1 Sep 2014 - 31 Aug 2015

4. TITLE AND SUBTITLE

Early Detection of NSCLC Using Stromal Markers in Peripheral Blood

5a. CONTRACT NUMBER
W81XWH-14-1-0263

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

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Dingcheng Gao, PhD
Nasser Altorki, MD
Olivier Elemento, PhD

E-Mail: dig2009@med.cornell.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

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

Joan & Sanford I. Weill Medical College
of Cornell University
1300 York Avenue, Box 89
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

There is an immediate clinical need to develop a diagnostic biomarker for lung cancer in early stage. In this proposal, instead of tumor-derived biomarkers, we are focused 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. We proposed to identify the lung cancer-dependent transcriptomic signature by analyzing flow cytometry purified circulating myeloid subpopulations from pre- and post-surgery lung cancer patients. During this research period, we have recruited 23 NSCLC patients and collected their peripheral blood before and after the surgical removal of the primary lung tumor. We have finished the data analysis of 17 of these patients. We have re-confirmed our approach to purify CD11b+CD33- and CD11b+CD33+ myeloid subpopulations by flow cytometry. The RNA-sequencing results are promising. A preliminary lung cancer specific gene signature was identified. We will continue our research by finishing the entire RNA-sequencing data from all patients and move forward to validate its diagnostic value for early stage lung cancer in the next research period.

15. SUBJECT TERMS
Non-small cell lung cancer (NSCLC), biomarker, circulating myeloid cells, flow cytometry, RNA-sequencing, expression profiling.

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	11	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

Page

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:**

Non-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 proposed to accomplish the Specific Aim 1 related tasks.

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. We have finished the data analysis of 17 of these patients. Their basic clinic information has been summarized in the Table 1. Of note, the majority of these NSCLC patients are Stage I or II (15 out of 17). To eliminate the effect of surgery on circulating myeloid cells, we have collected the post-surgical blood at least 4 weeks after the surgery. The samples (both pre and post-surgery) from rest of patients have also been submitted for flowcytomety, RNA-sequencing and in data-analyzing stage currently.

Table 1. Recruited NSCLC patients

NSCLC	(*n* = 17)
Age (y)	
Average	69
Max	84
Min	49
Gender	
Female	11
Male	6
Tobacco use	
Current	2
Former	13
Never	2
Cancer stage	
Stage I	14
Stage II	1
Stage III	2
Surgery to post-collection	
1-2 mo	12
2-4 mo	3
4-7 mo	2

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). Consistent as we found in the preliminary study, the absolute number and the percentage of myeloid subpopulations showed broad variations between patients. This has further justified the usage of flow cytometry sorting to isolate subpopulations of myeloid cells to eliminate the inter-individual variations in cell numbers.

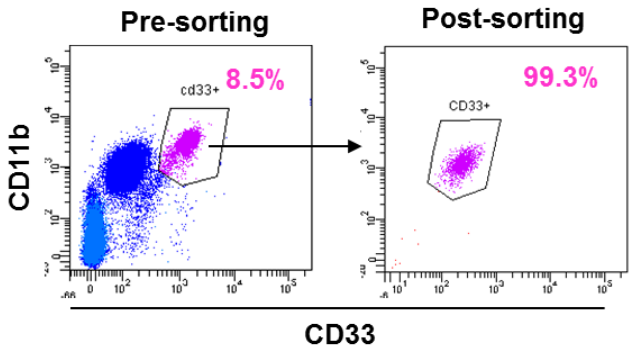

Figure 1. The purity of CD11b+CD33+ myeloid cells in the pre- and post-sorting samples.

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. Approximately 95% of total sequences were mapped to known RNA sequence in human genome, indicating high performance and data quality of RNA-seq analysis (Fig. 2).

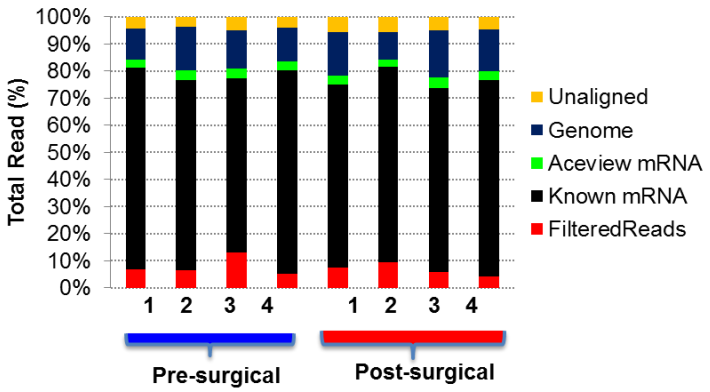

Figure 2. The alignment of RNA-sequencing result. RNA-sequencing results were aligned to human genome, the percentage of alignment indicated the performance of RNA-sequencing procedure.

Subtask 4: RNA-seq data analysis.

With the RNA-sequencing results from 17 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. 3). These results indicated the high reliability and reproducibility of expression profiles from different myeloid subpopulation cells.

To optimize the candidate list of differentially expressed gene, we have select genes that are up- or down- regulated by >1.5 fold in the pre-surgery samples as compared with post-surgery controls (adjusted p value <0.05). In addition, to ensure candidate genes can be reproducibly detected and validated by subsequent RT-PCR analysis, we excluded genes that were expressed at low levels (FPKM value < 5).

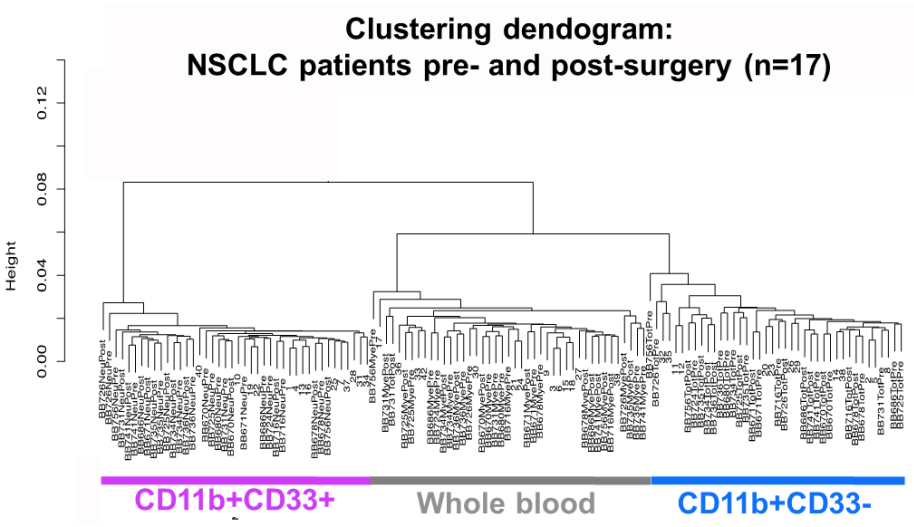

Figure 3. Clustering analysis of RNA-sequencing results of the sorted monocytes (CD11b+CD33+), neutrophils (CD11b+CD33-) and the unsorted whole blood. Differential gene expression profiles were detected from different subtypes of myeloid cells which were distinguishable from the.unsorted whole blood samples.

The analysis of the 17 paired pre- versus post-surgical samples has revealed that 22 and 203 genes from monocytes and neutrophils respectively, which were differentially expressed (Fig. 4A-B). We will finalize the lung cancer specific gene list as we get the analysis result from all 23 NSCLC patients during the next report period.

- **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 plan to continue the non-accomplished task related to the specific Aim 1. In the RNA-seq analysis, we will include the RNA-sequencing results of the pre- and post-surgical samples from all recruited 23 patients. We will finalized the lung cancer specific candidate gene list and design RT-PCR array for further analysis (Subtask 5: Feasible RT-PCR array assay development.)

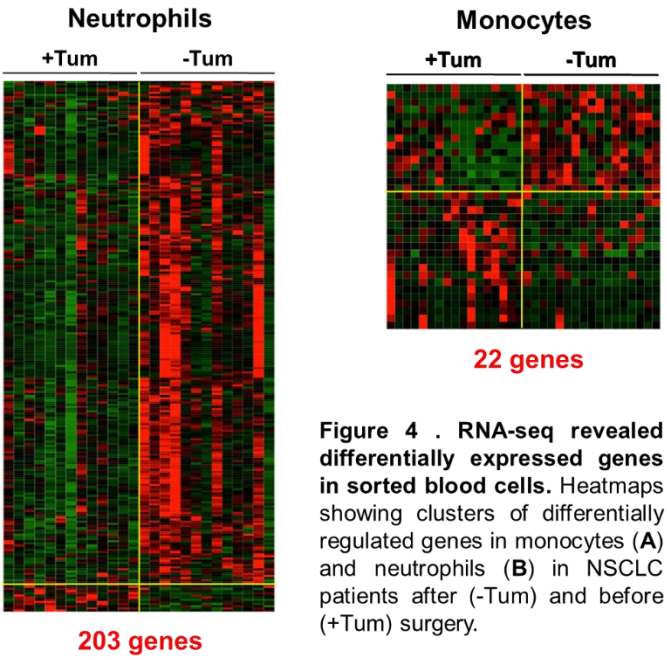

Figure 4 . RNA-seq revealed differentially expressed genes in sorted blood cells. Heatmaps showing clusters of differentially regulated genes in monocytes (**A**) and neutrophils (**B**) in NSCLC patients after (-Tum) and before (+Tum) surgery.

  We will then continue studies related to the Specific Aim 2: "To validate the diagnostic value of the specific gene signatures of circulating myeloid cells in patients with lung nodules." and accomplish the following tasks as proposed:

  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

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 learned that sorted circulating myeloid cells

possess unique expression signature, which may reveal the lung cancer specific biomarkers for early lung cancer detection.

- **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:**

   Nothing to Report

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:	*1.2 calendar months*
Contribution to Project:	*Dr. Gao has overseen the ongoing project, performed work in lung cancer biomarker discovery by combining flow cytometry and RNA-sequencing techiques.*
Funding Support:	NCI & Free to Breathe Foundation

Name:	*Nasser Altorki*
Project Role:	*Co-Investigator*
Researcher Identifier (e.g. ORCID ID):	

Nearest person month worked:	0.6 calendar months
Contribution to Project:	*Dr. Altorki has performed work in lung cancer biomarker discovery and has ensured relevant clinical samples are used. He has also supervised the biobanking of human samples and subsequent data analysis.*
Funding Support:	

Name:	*Olivier Elemento*
Project Role:	*Co-Investigator*
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	0.6 calendar months
Contribution to Project:	*Dr. Elemento has spearheaded the data analysis of RNA sequencing data that has emerged from the human samples.*
Funding Support:	NCI

Name:	*Anna Durrans*
Project Role:	*Research Specialist & Biobank Manager*
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	2.45 calendar months
Contribution to Project:	*Dr. Durrans has performed work in lung cancer biomarker discovery by combining flow cytometry and RNA-sequencing techniques.*
Funding Support:	AATS Foundation, Lung Cancer Research Foundation

Name:	*Jennifer Daniel*

Project Role:	*Research Technician & Biobank Manager*
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	1.9 calendar months
Contribution to Project:	*Ms. Daniel has performed work in lung cancer biomarker discovery by combining flow cytometry and RNA-sequencing techniques.*
Funding Support:	AATS Foundation

Name:	*Hyejin Choi*
Project Role:	*Post-Doc*
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	4.38 calendar months
Contribution to Project:	*Dr. Choi has performed work in lung cancer biomarker discovery by analyzing RNA-sequencing data and known crosstalk pathways between tumor and stroma.*
Funding Support:	NCI

Name:	*Cathy Spinelli, RN*
Project Role:	*Clinical Coordinator*
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	0.6 calendar months
Contribution to Project:	*Ms. Spinelli has ensured the appropriate human samples are collected and provided for the study and their associated clinical data.*

Funding Support:	NCI

Name:	*Abu Nasar, MS*
Project Role:	*Clinical Database Manager*
Researcher Identifier (e.g. ORCID ID):	
Nearest person month worked:	0.6 calendar months
Contribution to Project:	*Mr. Nasar has assisted in the analysis of clinically-relevant data pertaining to biomarker discovery,*
Funding Support:	NCI

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