

## Pathway-Based Approaches for Analysis of RNA-seq with SNPs-A Case Report Without Discovering Targeting Drugs for Metastatic Lung Cancer

Xiaonan Ying<sup>1</sup>, Hong Luo<sup>2</sup>, Yifei Zhang<sup>3</sup>, Jing Lu<sup>4</sup>, Wenqin Li<sup>5</sup>, Biaoru Li<sup>6\*</sup><sup>1</sup>Candidate for Master of Science in Bioinformatics, Northeastern University, Braintree, MA 02184, USA<sup>2,3</sup>Shanghai Xinchun Medical Tech. 588 Tian Xiong Rd, Shanghai, 201318, PRC<sup>4</sup>OBGYN, Monterey Park, CA 91754, USA<sup>5</sup>Department of Chemistry, University of California, Irvine, CA 92617, USA<sup>6</sup>Department of Pediatrics and GA Cancer Center Children Hospital at GA Augusta, GA 30912, U.S.A

\*Corresponding author: Biaoru Li

| Received: 18.01.2019 | Accepted: 28.01.2019 | Published: 28.02.2019

DOI: [10.21276/sjams.2019.7.2.6](https://doi.org/10.21276/sjams.2019.7.2.6)

## Abstract

## Original Research Article

Metastatic lung cancer is difficult to be cured due to easy and wide dissemination of the lung cancer. Here we reported a patient case suffering from small-cell lung cancer (SCLC) by using RNA-seq analysis. In the methodology, we used magnetic-activated cell sorting (MACS) to harvest a pair of the patient lung cancer cells and normal cells and then extracted RNAs for RNA-seq to achieve gene expression files and SNP profile from lung cancer cells and normal control cells. After uncovering gene expression profiles and SNP profile, genomic expression signature (GES) was used to map quantitative network and then GES combined with SNP (single-nucleotide polymorphism) profile was screened to discover driving genes and driving mutations. We successfully discovered that ADAM19, NID1, SLIT3 and SVEP1 abnormally drive EGFR pathway to cause lung cancer although finally no any drug was uncovered target ADAM19, NID1, SLIT3 and SVEP1 related to a personalized therapy. The case with that analysis tells us that pharmacogenomics scientists need develop more targeting genes related to drugs in high expressed EGFR pathway and physicians' scientists also require extending more methods such as epigenomics, microRNA and non-coding RNA by RNA detection techniques with their therapeutic targeting methods.

**Keywords:** magnetic-activated cell sorting (MACS), gene expression signature (GES), single-nucleotide polymorphism (SNP) signature, quantitative network, drug discovery, Betweenness Centrality (BC) and Degree Centrality (DC).

**Copyright © 2019:** This is an open-access article distributed under the terms of the Creative Commons Attribution license which permits unrestricted use, distribution, and reproduction in any medium for non-commercial use (NonCommercial, or CC-BY-NC) provided the original author and source are credited.

### INTRODUCTION

Personalized therapy is a new generation of treatment to be directly tailored for physicians to care individual patient relying on personal genomic profiles [1-4]. It is often called as "the right treatment for the right person at the right time." Here, a clinical sampling along with driving genes and mutations has been successfully discovered for personalized therapy by using clinical sampling from magnetic-activated cell sorting (MACS) to pick up lung cancer cells and normal cells, analyzing their genomic expression signature (GES) and single-nucleotide polymorphism profiles (SNP profiles). After GES/SNP signature database were combined with quantitative bioinformatics analysis, we have successfully discovered driving genes and mutation from EGFR pathway for the patient. The analysis indicated that clinical genomics can contribute personalized therapy, but we still have long way to go for individualized therapy including discovering more compounds of

targeting therapy and developing more methods such as epigenomics, microRNA and non-coding RNA with their therapeutic targeting methods.

### CASE REPORT AND METHODS

**Patient and specimen:** The patient is 60 years old male with dry cough containing blood sputum. Multiple metastatic masses were found with increase of FDG uptake in the inferior lobe of left lung by PET/CT and multiple bone destructions were discovered in ribs and iliac bones, cervical, thoracic and lumbar vertebra. Metastatic chest fluid obtained from patient suffering from SCLC was diagnosed according to conventional pathology. After Informed consent from the patient, the metastatic blood chest fluid was sent to our clinical laboratory. As our previous report, Ficoll gradient solution isolated the buffy coat containing CD326 cells [5] and a magnetic-activated cell sorting (MACS) further harvested a pair of lung cancer cells and normal cells from the metastatic chest fluid by negative

selection of CD45 Ab beads and positive selection for CD326 beads [6]. After CD326 cells were collected, CD326 Ab PE-labeling cells are used to identify CD326 cells under fluorescent microscopy for the cell purity.

**RNA-seq and QC analysis:** RNA was extracted from both cells in which methods shown as reported previously [7]. Each RNA with triple repeats was sent to Genomic Core in Kansas University Medical Center to perform High-seq for the genomics experiments. Galaxy platform was used for the short oligonucleotide analysis aligned to human transcript reference sequences from the human hg19 and the expression analysis at transcript levels by HISAT2 and differential analysis were performed by Cufflinks and Cuffdiff in the Galaxy platform as routine performance [8]. The transcripts expression level between the pair of the RNA samples was used as RPKM (reads per kilobase of transcript per million mapped reads). In order to define tumor cell driving mutation, we further analyzed SNP in the transcripts to uncover significant SNP based on the genotype calls from the paired tumor cells and normal cell samples [9].

**Discovery of signature GES and SNP:** After GES and functional SNP were uncovered; we combined targeting genes by both GES and significant SNPs so that signature SNPs profiles were uncovered for the driving genes related driving mutation which also was used to search drugs for specifically targeting the SCLC [10].

## RESULTS AND DISCUSSION

**Tumor cells purification and RNA extract with QC analysis:** After metastatic chest fluid obtained from patient, Ficoll solution isolated the buffer colt and MACS was used to purify CD326 cells. As Fig-1 shown, CD326 can harvest 86% as Fig-1B after negative-positive selection (which means excluding leukocytes by CD45 Ab beads) and positive collection by CD326 Ab beads under fluorescent microscopy while control cells only contained 4% CD326 cells as Fig-1A which was from negative selection workflow. The positive collection cells and control cells were used to RNA extract. RNA quantity and quality all analyzed from Agilent 2100 were matched to RNA-seq requirement.

**RNA-seq performance and bioinformatics analysis:** After Genomic Core in Kansas University Medical Center performed High-seq for the experiments, we used Galaxy platform to mine the RNA-seq data and QC analysis, HISAT2 sequenced reads aligned to human transcript reference sequences from the human hg19. The performance of gene/transcript expression were subject to Cufflinks and Cuffdiff in the Galaxy platform. The performance of SNPs was subject to BWA sequenced reads aligned to human hg19a with the performance of SNP by SAM and GATK platform in the Galaxy system. After we achieved RPKM (reads per kilobase of transcript per million mapped reads) from tumor cells to compare to normal control cells, we first studied gene expression characteristics according to fold level change (>2, >1.5, >1.25 and <2, <1.5 and <1.25), P value and Q- value (which an adjustable P value) as Table-1. We used to P and Q value lesser than 0.05 and fold change larger than 2-fold to uncover significant gene expression list.

We secondly uncovered significant SNP based on the genotype calls from the paired tumor cells and normal cell samples. As Table-2, we input 420,000 SNP which discovered from GATK platform so that we achieved 87525 SNPs from exon and splicing. After removed synonymous and non-frameshift SNP, we further study conserved SNP and harvest 2155 SNPs as significant SNP for next step work.

**Results of signature GES and SNP for driving genes and mutations with their drugs search:** After GES and functional SNP were uncovered, we observed that 552 transcripts (with P and Q values <0.05 and ontology distribution as Fig-2) and combined 2155 SNPs were further performed tumor driving genes. A list of driving genes was discovered as Table-2 from both GES and significant SNPs. We discovered ADAM19, NID1, SLIT3 and SVEP1 related to EGF pathway. We also discovered COMT, CHRNB4, IL7R and NCF2, which are increasingly reported to relate with lung cancer. Although we successfully uncover targeting genes (ADAM19, NID1, SLIT3, SVEP1 related to EGF pathway and COMT, CHRNB4, IL7R and NCF2 highly expressed in lung cancer), current target drugs approved by FDA cannot be discovered from these driving genes.

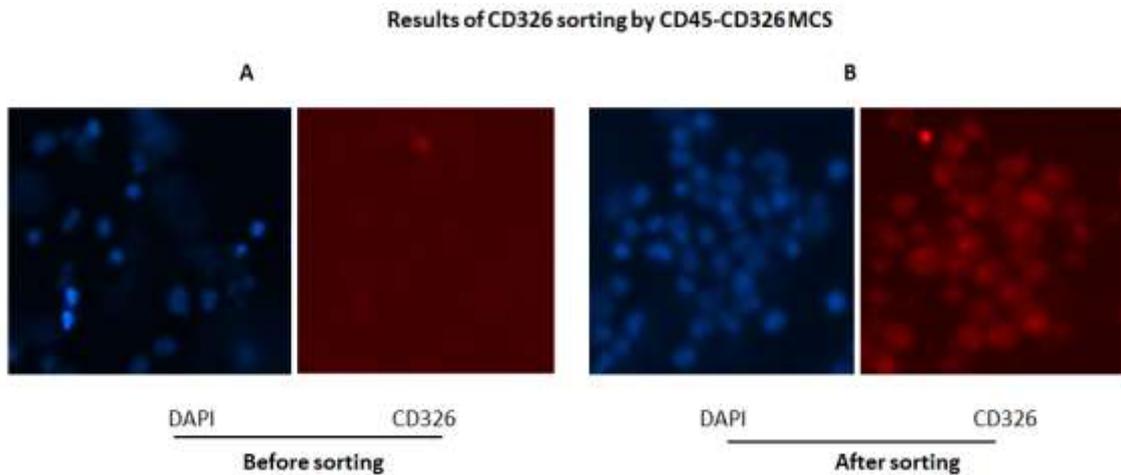
**Table-1: Results of Gene Expression profiles**

	>2 fold	>1.5 fold	> 1.25fold	<2 fold	<1.5 fold	<1.25 fold
Fold Change	2213	3987	6569	4908	7979	11188
P value	314	722	1543	366	1602	3167
Q value	314	722	1543	351	1531	2895
Defined gene	243	547	1228	155	708	1369

**Table-2: Results of SCLC driving genes (A and B)**

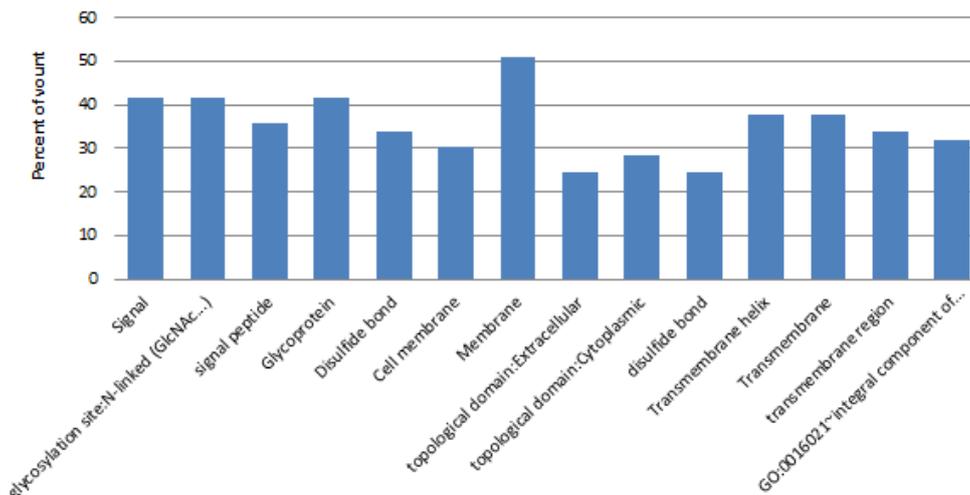
ID (EGF pathway) A	Gene Name	Species
ADAM19	ADAM metallopeptidase domain 19(ADAM19)	Homo sapiens
NID1	nidogen 1(NID1)	Homo sapiens
SLIT3	slit guidance ligand 3(SLIT3)	Homo sapiens
SVEP1	sushi, von Willebrand factor type A, EGF and pentraxin domain containing 1(SVEP1)	Homo sapiens
ID (Lung Cancer) B	Gene Name	Species
COMT	catechol-O-methyltransferase(COMT)	Homo sapiens
CHRN4	cholinergic receptor nicotinic beta 4 subunit(CHRN4)	Homo sapiens
IL7R	interleukin 7 receptor(IL7R)	Homo sapiens
NCF2	neutrophil cytosolic factor 2(NCF2)	Homo sapiens

**Legends**



**Fig-1: MACS performance for CD326 and CD45**

**Ontology of Significant Transcripts related Lung Cancer**



**Fig-2: The pathway and construction were defined by gene expression profile**

## CONCLUSION

Personalized therapy has increasingly reported to treat uncured tumor disease so that we have successfully reported several cases which have been successfully used by the new generation of treatment [11-13]. This report demonstrated a performance for personalized therapy by using analyzing their genomic expression signature (GES) and single-nucleotide polymorphism profiles (SNP profiles) to discover driving genes and mutation regarding EGFR pathway which can cause the lung cancer, although the target gene and mutation have not related to target drugs. The analysis tells us that scientists need develop more targeting genes related to drugs in EGFR related pathway to cure lung cancer and physicians' scientists require extending more methods such as epigenomics, microRNA and non-coding RNA by RNA detection techniques with their methods.

## Acknowledgments

Under the support of Dr. H. D. Preisler, we have set up different methods and models to analyze genomic profiles such as CD3, CD4 and CD8 from immune and tumor diseases related personalized therapy. This clinical application was previously supported by National Cancer Institute IRG-91-022-09, USA (to BL).

## Authors contributions

XY and WL analyze topology and quantitative network under guidance of BL including python scripts; HL and FZ performed RNA under guidance BL; BL and J give clinical support to some drugs definition; BL conceived and designed the experiments.

## Competing interest statements

The authors declare no financial interests.

## REFERENCES

- Rodríguez-Antona C, Taron M. Pharmacogenomic biomarkers for personalized cancer treatment. *Journal of internal medicine*. 2015 Feb;277(2):201-17.
- Schilsky RL. Opinion: Personalized medicine in oncology: the future is now. *Nature Reviews Drug Discovery*. 2010; 9, 363-366.
- Hudson TJ. Genome variation and personalized cancer medicine. *J Intern Med*. 2013; 274(5):440-50.
- Yoshida T, Ono H, Kuchiba A, Saeki N, Sakamoto H. Genome-wide germline analyses on cancer susceptibility and GeMDBJ database: Gastric cancer as an example. *Cancer Sci*. 2010; 101(7):1582-9.
- Leij LD, Helrich W, Stein R, Mattes MJ. SCLC-cluster-2 antibodies detect the pancarcinoma/epithelial glycoprotein EGP-2. *Int J Cancer*. 1994; 8:S60-S63.
- Bantikassegn A, Song X, Politi K. Isolation of epithelial, endothelial, and immune cells from lungs of transgenic mice with oncogene-induced lung adenocarcinomas. *Is J Respir Cell Mol Biol*. 2015 52(4):409-17.
- Hu H, Zhang Q, Li S, Zhang X, Han J, Xiao H, Yan D, Zheng J, Li B. A Therapeutic Targeting Identification from Microarray Data and Quantitative Network Analysis. *The Open Access Journal of Science and Technology*. 2015 Mar 20;3.
- Chen SH, Kuo WY, Su SY, Chung WC, Ho JM, Lu HH, Lin CY. A gene profiling deconvolution approach to estimating immune cell composition from complex tissues. *BMC Bioinformatics*. 2018; 19(4):154.
- Nijveen H, van Kaauwen M, Esselink DG, Hoegen B, Vosman B. QualitySNPng: a user-friendly SNP detection and visualization tool. *Nucleic acids research*. 2013 Apr 30;41(W1):W587-90.
- Hobeika C, Rached G, Eid R, Haddad F, Chuciri S, Kourie HR, Kattan J. ALK-rearranged adenocarcinoma transformed to small-cell lung cancer: a new entity with specific prognosis and treatment?. *Personalized medicine*. 2018 Jan 30;15(2):111-5.
- Sotiriou C, Piccart MJ. Taking gene-expression profiling to the clinic: when will molecular signatures become relevant to patient care?. *Nature Reviews Cancer*. 2007 Jul;7(7):545.
- Patsialou A, Wang YR, Lin J. Selective gene-expression profiling of migratory tumor cells in vivo predicts clinical outcome in breast cancer patients. *Breast Cancer Res* 2012;14(5): R139.
- Zheng Y, Zhou J, Tong Y. Gene signatures of drug resistance predicts patient survival in colorectal cancer. *The pharmacogenomics journal*. 2015 Apr; 15(2):135.