Utility of Cell-Free DNA from Bronchoalveolar Lavage Fluids Using Next Generation Sequencing in Predicting Malignant Solitary Pulmonary Nodules

Jia-Chao Qi

Zhangzhou Municipal Hospital of Fujian Province and Zhangzhou Affiliated Hospital of Fujian Medical University

Li Lin

Zhangzhou Municipal Hospital of Fujian Province and Zhangzhou Affiliated Hospital of Fujian Medical University

Zhiwei Zhao

Zhangzhou Municipal Hospital of Fujian Province and Zhangzhou Affiliated Hospital of Fujian Medical University

Liping Liao

Zhangzhou Municipal Hospital of Fujian Province and Zhangzhou Affiliated Hospital of Fujian Medical University

Yixuan Lin

Zhangzhou Municipal Hospital of Fujian Province and Zhangzhou Affiliated Hospital of Fujian Medical University

Tiezhu Wang

Zhangzhou Municipal Hospital of Fujian Province and Zhangzhou Affiliated Hospital of Fujian Medical University

Miaofen Hu

Zhangzhou Municipal Hospital of Fujian Province and Zhangzhou Affiliated Hospital of Fujian Medical University

Zhi Wu

Zhangzhou Municipal Hospital of Fujian Province and Zhangzhou Affiliated Hospital of Fujian Medical University

Yangwu Ou

Zhangzhou Municipal Hospital of Fujian Province and Zhangzhou Affiliated Hospital of Fujian Medical University

Zhiming Cai

Zhangzhou Municipal Hospital of Fujian Province and Zhangzhou Affiliated Hospital of Fujian Medical University

Yuming Ye

Zhangzhou Municipal Hospital of Fujian Province and Zhangzhou Affiliated Hospital of Fujian Medical University

Qiyin Wu

Zhangzhou Municipal Hospital of Fujian Province and Zhangzhou Affiliated Hospital of Fujian Medical University

Qiaozhen Xu

Zhangzhou Municipal Hospital of Fujian Province and Zhangzhou Affiliated Hospital of Fujian Medical University

Weiliang Zhang

Zhangzhou Municipal Hospital of Fujian Province and Zhangzhou Affiliated Hospital of Fujian Medical University

Wensen Huang

Zhangzhou Municipal Hospital of Fujian Province and Zhangzhou Affiliated Hospital of Fujian Medical University

Hao Li (✉ 342171157@qq.com)

Zhangzhou Municipal Hospital of Fujian Province and Zhangzhou Affiliated Hospital of Fujian Medical University <https://orcid.org/0000-0001-9636-069X>

Research Article

Keywords: Bronchoalveolar lavage fluids, Solitary pulmonary nodule, Next-generation sequencing, Cell-free DNA, Early diagnosis

Posted Date: December 15th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1071910/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Purpose: To explore the utility of cell-free DNA (cfDNA) from bronchoalveolar lavage fluids (BALF) using next generation sequencing (NGS) in differentiating malignant tumors from benign solitary pulmonary nodule (SPN).

Methods: Between January 1st 2019 and January 1st 2021, 40 subjects undergoing computed tomography (CT) examination who were diagnosed with SPN, were prospectively enrolled at Zhangzhou Affiliated Hospital of Fujian Medical University (Zhangzhou, Fujian, China). And pathological diagnosis were finally confirmed from tissue specimens obtained by surgical resection. For each patient, the mutations of gene were analysed using NGS in both extraction of cfDNA isolated from BALF and tissue.

Results: Of 40 patients, 55% of them were diagnosed with lung adenocarcinoma, 20% with benign nodules, and 10% with small cell carcinoma. And patients with squamous carcinoma, adenosquamous carcinoma or large cell neuroendocrine carcinoma account for 5%, respectively. And 62.5% of malignant SPN (10/16) had at least one alteration. The most common alterations were TP53 (31.25%), followed by EGFR (18.75%), KRAS (12.5%), PIK3CA (6.25%), ERBB2 (6.25%), ALK (6.25%) and ROS1 (6.25%). Besides, there are some differences shown in the heatmap of gene mutation in the histologic sample. And there was a colse correlation between the mutations found in the tissue and BALF. For all 40 patients, the sensitivity, specificity, and concordance of BALF in predicting malignant nodules were 68.8%, 100%, and 75%, respectively.

Conclusions: By NGS, tumor-specific mutations of cfDNA from BALF may be beneficial to predicting malignant SPN, which may be taken into consideration for personalized cancer diagnosis.

Introduction

Lung cancer is generally detected at advanced, inoperable staging, which leads to the most common cause of cancer-associated deaths for both sexes[1]. Therefore, early identifications and interventions in lung cancer would be of primary importance, especially for those patients with asymptomatic and curable solitary pulmoary nodule (SPN) [2].

It is now accepted that fiberoptic bronchoscopy is the key diagnostic approach in cases of suspected lung cancer[3], and the diagnostic yield of cytologic analysis of bronchoalveolar lavage fluids (BALF) for peripheral lung cancer reaches 65% [4]. It is well known that circulating tumor cells in BALF may be sensitive to discriminate malignant SPN[5]. Before becoming clinically apparent, the detection of molecular indicators is beneficial to a decreased mortality[6]. Cell-free DNA(cfDNA), derived from tumors, is likely to express the entire genomic landscape. Indeed, next-generation sequencing (NGS) for the detection of cfDNA has also made great breakthroughs[7].

To our knowledge, there is little information available in literature about cfDNA from BALF using NGS on patients with SPN. Schmidt's group succeeded in isolating cfDNA from BALF and identifying alterations.

In particular, they focused on cancer patients and utilized polymerase chain reaction (PCR)/reverse transcriptase (RT)-PCR. Meanwhile, the positive incidence of gene mutation was unsatisfactory [8]. In addition, both surgically resected lung peripheral nodules and plasma DNA were investigated using NGS by Ye and his colleagues. Unfortunately, they suggested low concordance between tissue DNA and ctDNA mutations, so more attempts are required in optimizing the model[9]. Besides, Buttitta enrolled patients with lung adenocarcinoma who underwent surgical resection or biopsy and evaluated EGFR mutations from BALF[10]. Because of the important role in the pathogenesis of lung cancer, accurate detection of gene mutations such as EGFR and KRAS can be associated with prognosis[11, 12].

Thus, we aimed to explore lung tumour-associated alterations using NGS in cfDNA from BALF and test its ability to predict malignant SPN in early diagnosis of lung cancer.

Patients And Methods

Patient recruitment and sample collection

Between January 1st 2019 and January 1st 2021, 40 subjects undergoing computed tomography (CT) examination who were diagnosed with SPN (nodule size<1cm) were prospectively enrolled at Zhangzhou affiliated hospital of Fujian Medical University (Zhangzhou, Fujian, China). The scanned DICOM data was imported to computer via VBN software (DirectPath V1.02, Cybernet Systems), creating the target virtual bronchoscope bronchial images automatically, and the guidance pathway to lesions was established. Bronchoscope (Olympus BF-P260F, outer diameter in 4.0mm, working aperture in 2.0mm) was navigated to the target of the bronchi by the VBN system, pushing into the ultrasonic probe (UM-S20-20R, Olympus) to the corresponding segment and then explored to low echo ultrasound images. Afterward withdrew the ultrasonic probe slowly and measured the distance from the opening of segmental bronchus to area of the lesion indicated by ultrasound. Then, according to the measured distance, repeated observing whether the operation path was correct by the ultrasonic probe twice. Warm saline (0.9% NaCl) was injected in 20 ml aliquots through the working channel and harvested by pooling into sterile collection tubes (yield 15–35 ml). From each patient, 8-10 ml of BALF was collected and centrifuged at 2500g for 15 min to separate supernatants. BALF supernatant was used for the extraction of cfDNA. Within one month, patients underwent surgery after BALF and pathological types were finally confirmed by pathologists. For each patient, both the extraction of cfDNA isolated from BALF supernatants and tissue were analysed for mutation using NGS. And TNM classification was performed in accordance with 2020 NCCN index.

The study was approved by the ethical committee of Zhangzhou Affiliated Hospital of Fujian Medical University(ethics approval no. Zzsyy-2017-1116), and all patients provided informed written consent. All samples were tested in a centralized clinical testing center (Nanjing Geneseeq Technology Inc., Nanjing, China).

DNA extraction, target capture, and next-generation sequencing

According to the manufacturer's protocol, we purified CfDNA from BALF supernatants utilizing the QIAamp Circulating Nucleic Acid Kit (Qiagen, Hilden, Germany) and extracted Genomic DNA from BALF sediment applying the DNeasy Tissue Kit (Qiagen). Then, Genomic DNA was qualified using a Nanodrop2000 (Thermo Fisher), and BALF-cfDNA fragment distribution was analyzed on a Bioanalyzer 2100 using the High Sensitivity DNA Kit (Agilent Technologies). We quantified all DNA by utilizing the dsDNA HS Assay Kit on a Qubit 3.0 Fluorometer (Life Technologies). Subsequently, we prepared sequencing libraries using the KAPA Hyper Prep Kit (KAPA Biosystems). Indexed DNA libraries were pooled together up to 2µg of total input and subjected to probe-based hybridization with GeneseeqOne NGS panel targeting 425 predefined cancer-associated genes (Geneseeq Prime panel). Captured libraries were amplified with Illumina p5 and p7 primers in KAPA HiFi HotStart ReadyMix (KAPA Biosystems). Finally, captured libraries were sequenced on Illumina HiSeq4000 NGS platforms (Illumina) with 2×150 bp pair-end reads.

Data processing and bioinformatics Analysis

We applied Trimmomatic for FASTQ file quality control[13]. Consequently, we removed leading/trailing low quality (quality reading below 15) or N bases. Then the data were aligned to the hg19 reference human genome with the Burrows-Wheeler Aligner (bwa-mem)[14]. According to the instruction, we applied local realignment around indels and base quality score recalibration with the Genome Analysis Toolkit (GATK 3.4.0) [15]. Normal and tumor BAM files were paired applying MuTect [16] with default parameters to identify somatic single nucleotide variants (SNVs). We analyzed small insertions and deletions (indels) using SCALPEL (<http://scalpel.sourceforge.net>). For BALF-sDNA, we required minimum variant allele frequency=1%, minimum variant supporting reads = 5. For BALF-cfDNA, we required minimum variant allele frequency= 1% or 0.3%, minimum variant supporting reads = 5 or 3, for non-hotspot and hotspot mutations (defined as recurrence≥20 in COSMIC database), respectively.

We excluded the SNVs and indels in the 1000 Genomes project and ExAC database with frequency>1%. Then, according to the hg19 reference genome and 2014 versions of standard databases and functional prediction programs, SNV and indel annotation was performed. Gene fusions were identified by FACTERA[17] and copy number variations (CNVs) were analyzed with ADTEX[18]. The log2 ratio cut-off for copy number gain was defined as 2.0 for BALF-sDNA and 1.6 for BALF-cfDNA samples. A log2 ratio cut-off of 0.6 was applied for copy number loss detection in all sample types.

Statistical analysis

Statistical analyses were conducted using SPSS 22.0 (Chicago, IL, USA). Categorical variables were described as number (percentage). They were compared applying the chi-square test or Fisher's exact test. Sensitivity, specificity were defined as follows: sensitivity= TP/ (TP+FN), specificity = TN/ (TN+FP); where TP is true positive, FN is false negative, TN is true negative, and FP is false positive. All p values described are two sided, and a p value of <0.05 was thought statistically significant.

Results

Demographic and clinical characteristics of the study cohort

Table 1 shows the demographic and clinical characteristics of all subjects. We included 24 men and 16 women, and the median age was 60 years (range 39-72 years). Of 40 patients, 55% of them were diagnosed with lung adenocarcinoma, 20% with benign nodules, and 10% with small cell carcinoma. And patients with squamous carcinoma, adenosquamous carcinoma or large cell neuroendocrine carcinoma account for 5%, respectively. The pathologic stages of all patients were stage I.

Table 1
Demographic and clinical characteristics of the study cohort

	No of patients
Gender	
Male	24(60%)
Female	16(40%)
Clinical stage (TNM staging)	
I	40(100%)
Smoking status	
Never	26(65%)
Ever	14(35%)
Pathological diagnosis	
Adenocarcinoma	22(55%)
Squamous carcinoma	2(5%)
Adenosquamous carcinoma	2(5%)
Small cell carcinoma	4(10%)
Large cell neuroendocrine carcinoma	2(5%)
Benign nodule	8(20%)
Data are presented as number (%)	

Description Of Alterations And Actionability Of The Detected Alterations

Table 2 shows the pathological diagnosis and mutation type of the study cohort. Most malignant nodule had realated driver genes. And detailed distributions were described in the heatmap of gene mutation in the BALF sample (Fig. 1). Patients had a median of one alteration (range, 0–10), and 62.5% of malignant patients (20/32) had at least one alteration. The most common alterations in BALF were TP53 (31.25%), followed by EGFR (18.75%), KRAS (12.5%), PIK3CA (6.25%), ERBB2 (6.25%), ALK (6.25%), NTRK (6.25%) and ROS1 (6.25%). TP53 and RB1 are the top two mutated tumor suppressor genes, with a frequence of 31.25% and 12.5%, respectively (Fig. 1). Other top mutated genes include tumor suppressor genes FAT1 (6.25%), as well as ARID1A (6.25%), KMT2B (6.25%). In addition, there are some differences shown in the heatmap of gene mutation in the histologic sample (Fig. 2), such as ALK, ROS1, EGFR and TP53.

Table 2
Pathological diagnosis and mutation of BALF of the study cohort

Number of Patient	Malignant diagnosis	Pathological Type	Mutation of BALF	Driver Genes
01	Yes	Adenocarcinoma	Yes	EGFR 19 Del
02	Yes	Adenocarcinoma	Yes	EGFR L858R
03	No	Benign nodule	No	-
04	No	Benign nodule	No	-
05	Yes	Adenocarcinoma	No	ROS1
06	Yes	Small cell carcinoma	Yes	NTRK1
07	No	Benign nodule	No	-
08	Yes	Small cell carcinoma	Yes	-
09	Yes	Adenocarcinoma	Yes	EGFR 19 Del
10	No	Benign nodule	No	-
11	Yes	Adenocarcinoma	Yes	PIK3CA
12	Yes	Adenocarcinoma	Yes	-
13	Yes	Adenocarcinoma	Yes	KRAS
14	Yes	Adenocarcinoma	No	-
15	Yes	Adenosquamous carcinoma	Yes	ERBB2
16	Yes	Adenocarcinoma	Yes	KRAS/ALK
17	Yes	Large cell neuroendocrine carcinoma	No	-
18	Yes	Adenocarcinoma	Yes	-
19	Yes	Adenocarcinoma	No	-
20	Yes	Squamous carcinoma	Yes	-
21	Yes	Adenocarcinoma	Yes	EGFR 19 Del
22	Yes	Adenocarcinoma	Yes	EGFR L858R

BALF:bronchoalveolar lavage fluids

Number of Patient	Malignant diagnosis	Pathological Type	Mutation of BALF	Driver Genes
23	No	Adenocarcinoma	Yes	KRAS/ALK
24	No	Adenocarcinoma	Yes	EGFR 19 Del
25	Yes	Adenocarcinoma	Yes	ROS1
26	Yes	Small cell carcinoma	No	-
27	No	Benign nodule	No	-
28	Yes	Small cell carcinoma	Yes	-
29	Yes	Adenocarcinoma	Yes	EGFR 19 Del
30	No	Benign nodule	No	-
31	Yes	Adenocarcinoma	Yes	PIK3CA
32	Yes	Adenocarcinoma	Yes	EGFR L858R
33	Yes	Adenocarcinoma	No	-
34	Yes	Adenocarcinoma	Yes	PIK3CA
35	Yes	Adenosquamous carcinoma	No	-
36	Yes	Adenocarcinoma	Yes	KRAS/ALK
37	Yes	Large cell neuroendocrine carcinoma	No	-
38	Yes	Adenocarcinoma	Yes	ERBB2
39	Yes	Adenocarcinoma	Yes	KRAS/ALK
40	Yes	Squamous carcinoma	Yes	-
BALF:bronchoalveolar lavage fluids				

Sensitivity, specificity, and consistency of predicting benign and malignant nodules with BALF

There was a colse correlation between the mutations found in the histologic sample and BALF. (Tab. 3). Also, a relatively colse association existed between the incidence of the malignant nodules and positive

mutation from BALF (Tab. 4). Among all 40 patients, the sensitivity, specificity, and concordance of predicting malignant nodules with BALF were 68.8%, 100%, and 75%, respectively.

Table 3
Concordance of mutation between histologic sample and BALF sample

All patients	Histology mutation (+)	Histology mutation (-)	Total	χ2	P value
BALF Mutation (+)	22	6	28	13.41	0.289
BALF Mutation (-)	2	10	12		
Total	24	16	40		
BALF:bronchoalveolar lavage fluids					

Table 4
Concordance of BALF in predicting benign and malignant nodules for all patients

All patients	Malignant nodules	Benign nodules	Total	χ^2	P value
Mutation (+)	24	0	24	2.35	0.126
Mutation (-)	10	6	16		
Total	34	6	40		
BALF:bronchoalveolar lavage fluids					

Discussion

NGS was performed on cfDNA derived from the BALF in 40 patients with SPN. The most frequent mutations are TP53, followed by EGFR, KRAS, PIK3CA, ERBB2, ALK, NTRK and ROS1, and it is notable that the positive rates of EGFR, KRAS and TP53 are significantly different from previous reports[19, 20]. Particularly, the positive diagnoses of KRAS and TP53 mutations appeared not to be associated with histologic subtypes. Further, the incidence of mutation in BALF sample was not inferior to that of

histological issue. Overall, our research may provide an clinically important method using BALF to predict malignant nodule in suspected cancer patients.

Indeed, the difference in the positives of each driver gene might be attributed to sample size and source, clinical histological type and sequencing panel. The sensitivity of the bronchoscopic method in peripheral tumor may be subjected to tumor size and location[21]. Clinicians should accurately locate the SPN and obtain effective BALF. A non-invasive approach to predict the malignancy of surgery-candidate SPN is urgently needed. So far, there is little data about clear relations between genetic alterations and tumour subtypes from BALF. Limited by small sample size, it is objectively thought that predicting malignant nodules through driver mutation detection based on cfDNA may need more cautions. Still, different from other samples reporting about a high incidence of KRAS in adenocarcinomas[22] and a significantly low incidence of TP53 in squamous cell carcinoma[23], we demonstrated a relatively close link between malignant nodules and mutations of KRAS and TP53 from BALF.

Although some mutations which are not specific to histologic type occur in NSCLC patients with an equal frequency in our findings. The conclusions still suggested that these genetic changes may be initial events leading to lung cancer development. EGFR and KRAS mutations may existed in some synchronous lesions[24]. In agreement with previous reports[25, 26], our data showed that L858R and 19del were the most common mutations. It might be associated with the fact that, EGFR mutations are more prevalent in NSCLC in Asia than in western countries. Different from previous researchers focusing on confirmed lung cancer patients[8–10], we speculated that early detection of diseases-related genes by NGS would be alarm to those people with SPN at high-risk lung cancer. Meanwhile, we hypothesized that for specific patients, tumor may have possible selection mechanisms around certain genes or pathways that are important for carcinogenesis.

It is worth mentioning that NGS panel with targeting 425 predefined cancer-associated genes was performed, and some rare mutations (such as NTRK, ARID1A, ARID2 and SETD2) were observed. As potential therapeutic targets including NSCLC and sarcomas[27], we highlight the need to routinely detect NTRK fusions to broaden the therapeutic options. Considering that tumor is normally a mixture of different cell types, inactivation mutations in several switch/sucrose non-fermenting subunits, such as ARID1A and ARID2, are identified in a significant proportion of lung tumors[28]. Besides, SETD2 as a potent tumor suppressor in lung adenocarcinoma, developed model systems to improve chromatin deregulation in lung cancer[29].

Similar with previous findings[20, 30], close correlation existed between the mutations found in the histologic sample and BALF in our study. We suggested that it may provide a useful means to predict malignant nodule by BALF. For those patients afraid of surgery without definite malignant imaging manifestations, BALF enable patients to be tested less invasively[31]. We thought that comprehensive analysis of gene mutations in BALF may be a helpful supplement to enhance the diagnostic yield of differentiating malignant SPN. Unlike other traditional biopsies, cfDNA could provide an overview of all the mutations, allowing for a more targeted treatment. Particularly, personalized prediction may be

recommended for those patients with negative mutation results to avoid unnecessary surgery, which enable clinicians to more accurately judge treatment options.

The present findings, however, should be interpreted taking into account some limitations. First, the sample size was too small to perform additional statistical analyses on factors such as nodule location and nodule size, that could affect the yield of procedure. However, in line with previous observations[8, 32], our findings illustrated relatively high positive rates of gene mutations, suggesting sufficient DNA amplified by NGS from cell-free lavage supernatants. Of note, BALF needs less caution for specimen handling and is expected to minimize the tumor heterogeneity because of release of cfDNA fragments. Secondly, considering the cost, the patients did not utilize NGS on serum specimens, it may affect the correlation analysis between the mutations found in the serum sample and the BALF to some extent. Nevertheless, the mutations in pathological tissue and BALF maintained good agreement. Thirdly, it is noteworthy that KRAS mutation in BALF was detected not only from patients with lung cancer, but also from patients with benign lung disease[33]. Also, KRAS could be linked with smoking or chronic inflammatory processes[23].

Although limited by the sensitivity, tumor-specific mutations of cfDNA from BALF specimens using NGS would be beneficial to predicting malignant SPN. Cancer early detection is by far the most economical and effective mean to reduce cancer-specific mortality. And comprehensive targeted NGS enhances personalized cancer treatment.

Declarations

Funding: This work was funded by Startup Fund for scientific research, Fujian Medical University (grant 2017XQ1116), Natural Science Foundation of Fujian Province (grant 2020J01122220) and Youth Research Fund from Fujian Provincial Health Bureau (grant 2017-1-87).

Conflicts of interest: Jia-Chao Qi has received research grants from grant 2017XQ1116 for Startup Fund for scientific research, Fujian Medical University and grant 2017-1-87 for Youth Research Fund from Fujian Provincial Health Bureau. And Li Lin has received research grant 2020J01122220 Natural Science Foundation of Fujian Province. The authors had no relevant financial or non-financial interests to disclose. Jia-Chao Qi and Li Lin contributed equally to this work.

Ethics approval: All procedures performed in the studie involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.The study was approved by the ethical committee of Zhangzhou Affiliated Hospital of Fujian Medical University.

Informed consent: Informed consent was obtained from all individual participants included in the study.

Consent to publish: The authors affirm that human research participants provided informed consent for publication.

Authors' contributions: Hao Li conceived the study. Jia-Chao Qi, Li Lin and Zhi-Wei Zhao designed and performed the experiment; analyzed the data; and contributed to the manuscript preparation. Li-Ping Liao, Yi-Xuan Lin, Tie-Zhu Wang and Miao-Fen Hu contributed to the design of the study and analyzed the data. Zhi Wu, Yang-Wu Ou, Zhi-Ming Cai, Yu-Ming Ye, Qi-Yin Wu contributed to the revision of the manuscript. Qiao-Zhen Xu, Wei-Liang Zhang, Wen Sen Huang conducted the experiments. Jia-Chao Qi and Li Lin contributed equally to this work. All authors read and approved the final manuscript.

References

1. Siegel RL, Miller KD, Jemal A(2017) Cancer Statistics, 2017. *CA Cancer J Clin*;67: 7-30.
2. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D(2011) Global cancer statistics. *CA Cancer J Clin*;61: 69-90.
3. Karahalli E, Uuml, Yilmaz A, uuml, Rker H, Ouml, et al.(2001) Usefulness of various diagnostic techniques during fiberoptic bronchoscopy for endoscopically visible lung cancer: should cytologic examinations be performed routinely? *Respiration*;68: 611-614.
4. Lee HS, Kwon SY, Kim DK, Yoon HI, Yim JJ(2007) Determinants of diagnostic bronchial washing in peripheral lung cancers. *International Journal of Tuberculosis & Lung Disease the Official Journal of the International Union Against Tuberculosis & Lung Disease*;11: 227-232.
5. Zhong CH, Tong D, Zhou ZQ, Su ZQ, Luo YL, Xing J, et al.(2018) Performance evaluation of detecting circulating tumor cells and tumor cells in bronchoalveolar lavage fluid in diagnosis of peripheral lung cancer. *J Thorac Dis*;10: S830-S837.
6. Hirsch FR, Merrick DT, Franklin WA(2002) Role of biomarkers for early detection of lung cancer and chemoprevention. *European Respiratory Journal*;19: 1151-1158.
7. Abbosh C, Birkbak NJ, Wilson GA, Jamal-Hanjani M, Constantin T, Salari R, et al.(2017) Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. *Nature*;545: 446-451.
8. Schmidt B, Carstensen T, Engel E, Jandrig B, Witt C, Fleischhacker M(2004) Detection of cell-free nucleic acids in bronchial lavage fluid supernatants from patients with lung cancer. *Eur J Cancer*;40: 452-460.
9. Ye M, Li S, Huang W, Wang C, Liu L, Liu J, et al.(2018) Comprehensive targeted super-deep next generation sequencing enhances differential diagnosis of solitary pulmonary nodules. *J Thorac Dis*;10: S820-S829.
10. Buttitta F, Felicioni L, Del Grammastro M, Filice G, Di Lorito A, Malatesta S, et al.(2013) Effective assessment of egfr mutation status in bronchoalveolar lavage and pleural fluids by next-generation sequencing. *Clin Cancer Res*;19: 691-698.
11. Iwona O, Joanna D-K(2013) [Bronchoalveolar lavage in lung cancer–diagnostic value and assessment of the anti-cancer immune response]. *Postepy Hig Med Dosw:online*;67: 1119-1127.
12. Lang SM, Stratakis DF, Freudling A, Ebel K, Huber RM(2000) Detection of K-ras and p53 mutations in bronchoscopically obtained malignant and non-malignant tissue from patients with non-small cell

- lung cancer. *European Journal of Medical Research*;5: 341.
13. Bolger AM, Lohse M, Usadel B(2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*;30: 2114-2120.
 14. Li H, Durbin R(2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*;25: 1754-1760.
 15. Depristo MA, Banks E, Poplin R, Garimella KV, Daly MJ(2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*;43: 491-498.
 16. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al.(2013) Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature Biotechnology*;31: 213-219.
 17. Newman AM, Bratman SV, Stehr H, Lee LJ, Liu CL, Diehn M, et al.(2014) FACTERA: a practical method for the discovery of genomic rearrangements at breakpoint resolution. *Bioinformatics*;30: 3390-3393.
 18. Amarasinghe KC, Li J, Halgamuge SK(2013) CoNVEX: copy number variation estimation in exome sequencing data using HMM. *BMC Bioinformatics*;14 Suppl 2: S2.
 19. Wang S, Wang Z(2014) EGFR mutations in patients with non-small cell lung cancer from mainland China and their relationships with clinicopathological features: a meta-analysis. *Int J Clin Exp Med*;7: 1967-1978.
 20. Li J, Hu YM, Wang Y, Tang XP, Shi WL, Du YJ(2014) Gene mutation analysis in non-small cell lung cancer patients using bronchoalveolar lavage fluid and tumor tissue as diagnostic markers. *Int J Biol Markers*;29: e328-336.
 21. Schreiber G, McCrory DC(2003) Performance characteristics of different modalities for diagnosis of suspected lung cancer: summary of published evidence. *Chest*;123: 115s-128s.
 22. Keohavong P, Demichele MAA, Melacrinis AC, Landreneau RJ, Siegfried JM(1996) Detection of K-ras mutations in lung carcinomas: relationship to prognosis. *Clinical Cancer Research*;2: 411-418.
 23. Le Calvez F, Mukeria A, Hunt JD, Kelm O, Hung RJ, Tanière P, et al.(2005) TP53 and KRAS mutation load and types in lung cancers in relation to tobacco smoke: distinct patterns in never, former, and current smokers. *Cancer Res*;65: 5076-5083.
 24. Chung JH, Choe G, Jheon S, Sung SW, Kim TJ, Lee KW, et al.(2009) Epidermal growth factor receptor mutation and pathologic-radiologic correlation between multiple lung nodules with ground-glass opacity differentiates multicentric origin from intrapulmonary spread. *J Thorac Oncol*;4: 1490-1495.
 25. Ahn MJ, Park BB, Ahn JS, Kim SW, Kim HT, Lee JS, et al.(2008) Are there any ethnic differences in molecular predictors of erlotinib efficacy in advanced non-small cell lung cancer? *Clin Cancer Res*;14: 3860-3866.
 26. Huang W, Yin C, Wang G, Rosenblum J, Krishnan S, Dimitrova N, et al.(2019) Optimizing a Metatranscriptomic Next-Generation Sequencing Protocol for Bronchoalveolar Lavage Diagnostics. *J Mol Diagn*;21: 251-261.

27. Popper HH(2016) Commentary on tumor heterogeneity. *Translational Lung Cancer Research*;5: 433-435.
28. Walter DM, Venancio OS, Buza EL, Tobias JW, Deshpande C, Gudiel AA, et al.(2017) Systematic In Vivo Inactivation of Chromatin-Regulating Enzymes Identifies Setd2 as a Potent Tumor Suppressor in Lung Adenocarcinoma. *Cancer Res*;77: 1719-1729.
29. Majewski IJ, Nuciforo P, Mittempergher L, Bosma AJ, Eidtmann H, Holmes E, et al.(2015) PIK3CA mutations are associated with decreased benefit to neoadjuvant human epidermal growth factor receptor 2-targeted therapies in breast cancer. *J Clin Oncol*;33: 1334-1339.
30. Park S, Hur JY, Lee KY, Lee JC, Rho JK, Shin SH, et al.(2017) Assessment of EGFR mutation status using cell-free DNA from bronchoalveolar lavage fluid. *Clin Chem Lab Med*;55: 1489-1495.
31. Murtaza M, Dawson SJ, Tsui DW, Gale D, Forshew T, Piskorz AM, et al.(2013) Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. *Nature*;497: 108-112.
32. Schmidt B, Engel E, Carstensen T, Weickmann S, John M, Witt C, et al.(2005) Quantification of free RNA in serum and bronchial lavage: a new diagnostic tool in lung cancer detection? *Lung Cancer*;48: 145-147.
33. Oshita F, Nomura I, Yamada K, Kato Y, Noda K(1999) Detection of K-ras mutations of bronchoalveolar lavage fluid cells aids the diagnosis of lung cancer in small pulmonary lesions. *Clinical Cancer Research An Official Journal of the American Association for Cancer Research*;5: 617-620.

Figures

BALF:bronchoalveolar lavage fluids

Figure 1

Heatmap of gene mutation in the BALF sample

Figure 2

Heatmap of gene mutation in the histologic sample