RNF213 gene mutation of circulating tumor DNA in early diagnosis of NSCLC using targeted next-generation sequencing

Ning Jiang
Second Hospital of Shandong University
Shukang Zhao
Shandong University Cheeloo College of Medicine
Peichao Li
Shandong University Cheeloo College of Medicine
Yu Liu
PLA 960th Hospital
Hubo Shi
Shandong Provincial Chest Hospital
Chengke Zhang
Second Hospital of Shandong University
Yunshan Wang
Second Hospital of Shandong University
Chengjun Zhou
Second Hospital of Shandong University
Wenhao Zhang
Shandong University Cheeloo College of Medicine
Chuanliang Peng
Second Hospital of Shandong University
Weiquan Zhang
Second Hospital of Shandong University
Yingtao Hao
Second Hospital of Shandong University
Yunpeng Zhao
Second Hospital of Shandong University
Qifeng Sun
Shandong Provincial Hospital
Xiaogang Zhao (Menglin2002@163.com)

Research article

Keywords: Lung cancer; Early diagnosis; Targeted next generation sequencing (NGS); RNF213 gene; Circulating tumor DNA (ctDNA)

Posted Date: July 16th, 2019

DOI: https://doi.org/10.21203/rs.2.11443/v1

Abstract

Introduction To distinguish early stage lung cancer from benign disease of the lung nodules, especially the lesions with ground-glass opacity (GGO) or ground-glass nodule (GGN), we assessed gene mutations of the ctDNA in peripheral blood by using targeted next-generation sequencing (NGS). Methods Single lung nodule patients without mediastinal lymph nodes or symptoms hardly diagnosed by chest CT and biomarker of lung cancer were enrolled. All patients received minimally invasive surgery but refused preoperative biopsy. Gene mutations of pre-operative blood samples were detected by targeted NGS. Mutations with statistical differences were screened in lung cancer and benign disease grouped by postoperative pathology. Gene expression was determined by immunohistochemistry. Highly expressed genes were selected as biomarkers to verify the mutations in peripheral blood. Results In training set, RNF213, KMT2D, CSMD3 and LRP1B genes mutated more frequently in early stage lung cancer (25cases) than benign nodules (18cases) ($P<0.05$). High expressions of RNF213 gene in lung cancers and low expressions in benign diseases were evaluated by immunohistochemistry. RNF213 gene mutated in 25% lung cancer samples in the validation set of 28 samples and showed high specificity (100%) and low sensibility (25.9%). In GGO and GGN patients, RNF213 mutated more frequently in early stage lung cancer compared to benign diseases ($P<0.05$). Conclusions RNF213 gene mutation was observed more frequently in early stage lung cancer, but rather than benign nodules. Mutation of RNF213 gene in peripheral blood may be a high specificity biomarker and valuable for early diagnosis of lung cancer.

Background

Lung cancer remains a life-threatening malignancy with the highest morbidity and mortality in the world [1]. Five-year survival of lung cancer patients is still low [2,3] despite the using of molecular diagnosis and targeted therapy currently. Early diagnosis and treatment are effective ways to improve the survival of lung cancer patients. Using low dose computed tomography (LDCT) in screening could reduce lung cancer related mortality, and smaller lung nodules could be found in early stage. But the diagnosis may be difficult in some cases with atypical CT imaging, and traditional biomarkers such as carcino-embryonic antigen (CEA), neuro-specific enolase (NSE) and cytokeratin 19 (CYFRA-211) could not be satisfied for early diagnosis. Aspiration biopsy or surgery may be needed in most patients to confirm whether the nodules are malignancy or benign disease.

The ideal diagnostic method should be simple, less traumatic, easy to obtain and high positive rate. Circulating cell-free DNA (cfDNA) is a fragment of DNA released through cell apoptosis widely existing in blood, cerebrospinal fluid, urine and saliva [4 , 5]. As cfDNA could also be released by tumor cells through apoptosis and necrosis [6, 7], this DNA is called circulating tumor DNA (ctDNA). Liquid biopsy of the blood ctDNA detection is important in the diagnosis, monitoring and prognosis of the tumor [8].

The patient's ctDNA is more meaningful to better understanding the disease. CtDNA reflecting the character of somatic genetic features of the primary tumor [9] can be detected in the peripheral blood of patients with advanced cancers, and be used for monitoring therapeutic effect [10, 11]. The content of

plasma ctDNA accounts for 0.01% of cfDNA [12]. Studies [13] have indicated that the concentration of ctDNA in the plasma increases with stage probably because of the increasing of tumor burden. Very low level of detectable ctDNA in plasma and unknown mutations limited the potential application in diagnosis of early stage lung cancer.

With the development of sensitivity of next-generation sequence (NGS), the low-level concentration of the ctDNA in blood can be detected. At present ctDNA of advanced stage lung cancer has been studied in blood for monitoring therapeutic effect. Fewer studies were aimed at early stage lung cancer by detecting tumor DNA in tissue, or identifying mutations in ctDNA for lung cancer patients with limited number of genes [14]. Some lung cancer-related genes such as EGFR, ALK, and KRAS were usually used for targeted NGS in early stage lung cancer [15, 16]. Only few genes from the panel were used for targeted NGS. In addition, healthy or benign nodule individuals need to be used as the control group. So far there was no study to address whether ctDNA can be detected in lung benign nodules or whether there are differences of ctDNA in undiagnostic lung nodules including early stage lung cancer and benign disease.

Here we study ctDNA through targeted NGS in small lung nodules that cannot be clearly diagnosed by chest CT. A panel of 560 tumor-related hot spot genes was used to evaluate the targeted sequencing for plasma ctDNA in malignant and benign lung nodules. We hope to find out discrepant ctDNA in the two groups to guide diagnosis in early stage lung cancer.

Methods

Patients

Patients with single lung nodules were diagnosed in 2017-2018 enrolled in the study came from the Second Hospital of Shandong University, Shandong Provincial Chest Hospital and The 960th Hospital of People's Liberation Army of China. Lung cancer or benign disease could not be confirmed in the chest CT. The largest diameter of the lesion was less than 5cm in diameter and there were no involvement of mediastinal lymph nodes in CT imaging. Clinical stage was less than T2N0M0 (stage II, TNM stage 7 edition) if the nodule was considered to be lung cancer. In terms of preoperative routine examination, there were no metastatic lesions and no patients with other oncology history. Lung cancer related biomarkers such as CEA, NSE and CYFRA-211 could not help making definite diagnosis in the patients. All patients refused biopsy or it was difficult to obtain tissues for histologic diagnosis. All patients accepted minimally invasive thoracoscopic surgery.

Study design

A training set was established. In accordance with uniform diagnostic criteria, inclusion criteria and exclusion criteria, 42 cases met the standard and passed the blood sample test in all 58 registered patients. Qualified paired samples were sequenced by targeted NGS with a panel of 560 tumor-related hot

spot mutant genes. Mutated ctDNA was analyzed in lung cancer and benign disease according to histopathological results. We selected significantly different ctDNA in lung cancer group compared to benign disease control group in the results. Immunohistochemical staining was performed in formalin fixed paraffin-embedded (FFPE) tissue samples of these patients to analyze the expression of the selected ctDNA. Finally, high expression of the selected ctDNA in lung cancer was confirmed. A validation set included unknown pathological lung nodules was established and sequenced by the same panel NGS to test the selected ctDNA mutations.

Blood sample preparation

10 ml peripheral blood was sampled 1-3 days before operation. Blood samples in EDTA tubes were centrifuged for 10 minutes at 1600g at 4℃ and white blood cells were collected and stored. The supernatants from these samples were further centrifuged at 16,000 g for 10 min at 4℃, and plasma was collected and stored at −80℃ until use. White blood cell DNA was isolated using the DNA Isolation Kit for Mammalian Blood (Roche) and cfDNA was isolated using the QIAmp Circulating Nucleic Acid Kit (QIAGEN) according to the manufacturer's protocol. 10-50ng cfDNA was acquired from 1 ml plasma.

Genomic DNA preparation and targeted sequencing

The quality of genomic DNA about degradation and contamination was monitored on 1% agarose gel, while the concentration was measured by Qubit® DNA Assay Kit in Qubit® 2.0 Fluorometer (Life Technologies, Carlsbad, CA).

We designed probes on the website of Agilent about particular genes according the design description to get the target gene regions. Briefly, 180-280bp fragments were produced from fragmentation carried out by hydrodynamic shearing system (Covaris, Massachusetts, USA). Extracted DNA was then amplified by ligation-mediated PCR (LM-PCR), purified, and hybridized to the probe for enrichment. Non-hybridized fragments were washed out subsequently. Both captured and non-captured LM-PCR products were subjected to real-time PCR to estimate the magnitude of enrichment. High-throughput sequencing was carried out at the average 1000× sequence depth when each captured library was loaded on an Illumina Hiseq4000 platform (Illumina, San Diego, California, USA). Each captured library was sequenced independently to ensure that each sample met the desired average fold coverage.

Sequence data quality control

The original fluorescence image files obtained from HiSeq platform were transformed to short reads (Raw data) by base calling and recorded in FASTQ format, which contained sequence information and

corresponding sequencing quality information. Clean reads were acquired by excluding reads containing adapter contamination and low-quality/unrecognizable nucleotides. Downstream bioinformatical analyses were based on these clean data. At the same time, the total reads number, sequencing error rate, percentage of reads with average quality >20 and with average quality >30, and GC content distribution were calculated.

Reads mapping and somatic genetic alteration detection

Valid sequencing data were mapped to the reference human genome (UCSC hg19) by Burrows-Wheeler Aligner (BWA) software to get the original mapping results stored in BAM format [17]. Then, SAM tools [18] and Picard (http://broadinstitute.github.io/picard/) were used to sort BAM files and do duplicate marking, local realignment, and base quality recalibration to generate final BAM file for computing the sequence coverage and depth.

MuTect and Strelka softwares [19, 20] respectively were used for calling somatic single nucleotide variations (SNVs) and small insertions and deletions (InDels) from paired tumor-normal samples. In addition to default filters, polymorphisms of somatic SNVs and InDels referenced in the 1000 Genomes Project[21] or Exome Aggregation Consortium (ExAC)[22] with a minor allele frequency over 1% were removed. Subsequently, VCF (Variant Call Format) was annotated by ANNOVAR software[23].

Immunohistochemical (IHC) analysis

Immunohistochemical (IHC) analysis was performed on 5 μm thick sections derived from formalin-fixed and paraffin-embedded lung cancer and benign disease tissue samples. In brief, all the slides were dewaxed with xylene and a graded ethanol series, antigen repaired in citrate buffer (Beyotime Institute of Biotechnology, Haimen, China), organization background closed with 1.5% goat serum (Beyotime Institute of Biotechnology, Shanghai, China), then incubated with the primary antibodies at 4 °C overnight. Primary antibodies against KMT2D (cat no. 27266-1-AP) and RNF213 (cat no. 21028-1-AP) were obtained from Proteintech Group (Wuhan, Hubei, China). Primary antibodies against LRP-1B (cat no. NBP2-49582) and CSMD3 (cat no. NBP1-86371) were purchased from Novus Biologicals (Centennial, CO, USA). Then, the slides were washed and stained with the secondary antibody (Goat anti-Rabbit IgG H&L (HRP), cat. no. ab205718; Abcam) and DAB disclosure, counterstained with hematoxylin, dehydrated and mounted. The results were evaluated independently by two independent pathologists.

Either membranous or cytoplasm staining for KMT2D, RNF213, LRP-1B and CSMD3 was defined as positive. The staining intensity and extent of the staining area were graded according to the semi-quantitative scoring system. Staining intensity was characterized as follows: 0, none; 1, weak; 2, intermediate; and 3, strong. The extent of staining was defined as: 0, none; 1, <1/100; 2, 1/100 to 1/10; 3,

1/10 to 1/3; 4, 1/3 to 2/3; and 5, >2/3 of cells expressing in the respective lesions[24].The final total score was achieved by adding the score of intensity and extent, which ranged from 0 to 8. Scores of 0 were defined as “-”, scores of 2-4 were defined as “+”, scores of 5-6 were defined as “++” and scores of 7-8 were defined as “+++”. Low expression was defined as scores of 0-4, and high expression was defined as scores of 5-8.

StatisticalAnalysis

BWA, Samblaster and Sambamba softwares were used for comparing the sequenced data with the reference genome. MuTect software was used to search for somatic single nucleotide variation (SNV) mutation. Strelka software was used to search for somatic insertion-deletion (InDel). ANNOVAR software was used for anotating the structure and function of the detected variation. Lung cancer and benign disease were divided into two groups. CtDNA somatic SNVs in the two groups were analyzed by chi-square test. The general character data were analyzed by student t test or one way ANOVA. Analyses were performed using the SPSS Statistics version 23 (IBM Corp). P value <0.05 was considered to be statistically significant difference.

Results

Patients’ general characteristics

In training set, we collected 58 pairs of blood samples. Samples of patients diagnosed with stage III lung cancer were excluded according to postoperative pathology. Rejecting contaminated genome and unqualified samples, a total of 42 pair samples were sequenced after being checked qualified. The patients’ general characteristics is shown in the table (Table 1).

There were 27 lung cancer and 15 benign disease patients in training set, in which there were 18 male and 9 female patients in lung cancer group, and 10 male and 5 female patients in benign disease group. The average age of the two groups were 63.1±8.9 and 52.3±10.7, respectively（P<0.05）. The number of stage I and stage II lung cancer patients were 19 (45.2%) and 8 (19.0%), respectively. Grouping by tumor size, there were 18 lung cancer patients (42.9%) and 11 benign desease patients (26.2%) in the group with a diameter less than 3 cm, while those with a diameter of 3 to 5 cmwere 9 (21.4%) and 4 (9.5%). There was no significantly statistical difference (P>0.05) in the general character between the two groups including gender, smoking history and tumor size. The data of two groups are comparable.

The pathology of lung cancer includes adenocarcinoma (18, 42.9%), squamous carcinoma (7, 16.7%) and other types (2, 4.8%) that were big cell carcinoma and sarcomatoid carcinoma. The histology of most lung cancer pateints was adenocarcinoma. In benign lung disease, the postopreative pathology was tuberculosis (3, 7.1%), inflammatory pseudotumor (9, 21.4%), hamartoma (2, 4.8%) and sclerosing hemangioma (1, 2.4%). The number of stage IA, IB, IIA and IIB lung cancer was 5 (11.9%), 14 (33.3%), 6

(14.3%) and 2 (4.8%), respectively. There was no significant difference in all general data (P<0.05) except in age .

Biomarkers detection

Blood biomarkers including CEA, NSE and CYFRA-211 were detected before surgery. In training set, the average vaules of these biomarkers were 3.04±1.64, 22.96±17.04, and 4.54±8.18 ng/ml in lung cancer, and 1.85±0.92, 20.20±7.02, and 1.66±0.86 ng/ml in benign disease, respectively, and there were no significant differences (P>0.05 for each comparison) in the two groups. In validation set, the average vaules of these biomarkers were 2.47±1.30, 14.23±4.90, 2.23±1.00 ng/ml in lung cancer and 1.72±1.02, 15.46±4.62, 2.18±0.82 ng/ml in benign disease groups, and there were no significant differences (P>0.05) in the two groups (Figure 1).

Positive staining of the biomarkers was judged according to the range of clinical reference values. In training set, there were 3/17 and 0/12 CEA positive, 8/16 and 6/12 NSE positive, and 2/13 and 0/12 CYFRA-211 positive patients respectively in lung cancer and benign disease group (P>0.05, respectively in three biomarkers). In validation set, there were 1/18 and 0/8 CEA positive, 4/18 and 1/8 NSE positive, and 3/18 and 1/8 CYFRA-211 positive patients respectively in lung cancer and benign disease group (P>0.05, respectively in three biomarkers) (Table 2).

Cell free DNA detected

Cell free DNA (cfDNA) has been detected in all of the 42 samples in training set. The concentration of cfDNA in lung cancer group and benign disease group respectively was 0.2-3.04 ng/μl and 0.21-1.25 ng/μl, average concentration respectively was 0.53±0.66 ng/μl and 0.54±0.29 ng/μl. The statistical differences was not significant in the two group (P>0.05).

Somatic mutation analysis

The number of mutated genes by targeted sequenced was total 246 in lung cancer and benign disease groups in training set. There were total 522 somatic mutations in the two groups including 374 somatic mutations dectected in lung cancer group and 148 mutations in benign disease group.(Figure 2)

Most of the mutations were single nucleotide variations (SNVs). In lung cancer group, there were 348 nonsynonymous mutations including 347 missense mutations and only one deletion mutation (InDel) in all 374 somatic mutations. In benign disease group, there were 136 nonsynonymous mutations (missense mutations) in all 148 somatic mutations. The average number of somatic mutation respectively was 13.85±7.25 and 9.87±4.27 in two group, and the missense mutation respectively was

12.85±7.08 and 9.07±3.84. The statistical differences was not significant between the two group (P>0.05). (Figure 2)

Analyzing the number of missense mutations in general characteristics of the two groups including the patients' gender, age, smoking history, tumor size, pathology, stage and N1 station lymphatic metastasis. There were no statistically significant differences in the comparison, respectively (P>0.05).(Figure 3)

CtDNA detected by targeted NGS

Sequenced 42 samples were divided into lung cancer and benign disease group according to postoperative pathology. There were 27 samples in lung cancer group and 15 samples in benign disease control group. Somatic mutations were detected in both two groups. In total 246 genes mutations were detected in the training set (Figure 4). The number of missense mutations of LRP1B, KMT2D, RNF213 and CSMD3 gene was 8 (8/27, 29.6%), 7 (7/27, 25.9%), 7 (7/27, 25.9%) and 6 (6/27, 22.2%) in lung cancer group, and none of these genes were detected in benign disease control group. The mutation sites of the four genes were shown in the Table 3. RNF213 gene mutations located at exon 17, 24, 26, 29 and 59 of chromosome 17. KMT2D gene mutations located at exon 4, 11, 16, 34, 38 and 41 of chromosome 12. CSMD3 gene mutations located at exon 10, 14, 20, 54 and 59 of chromosome 8. LRP1B gene mutations located at exon 2, 8, 16, 51, 83 and 89 of chromosome 2 (Table 3).

LRP1B, KMT2D, RNF213 and CSMD3 gene mutation in lung cancer were more than that in benign disease, and there were significant statistical differences between the two groups (P<0.05). The WHSC1 gene mutated in 10 samples of lung cancer and 8 samples of benign disease, respectively. The GDNF gene mutated in 7 samples in lung cancer and 1 samples in benign disease. There were less than 5 mutations of other genes detected in either one of the two groups. There were no significant statistical differences of these genes in two groups (P>0.05).

Immunohistochemical results

In training set, RNF213, LRP1B, KMT2D and CSMD3 genes were considered to be statistically significant difference in the analyzing of sequenced data in lung cancer compare to benign disease. Immunohistochemistry (IHC) was performed on specimens of 27 lung cancer and 14 benign disease FFPE tissues to detect the expression of RNF213, LRP1B, KMT2D and CSMD3. IHC was not carried out in one sample (B23) because the tissue was too little. After staining, taking photographs and evaluating, we show the representative illustrations of expressions of RNF213, KMT2D, CSMD3 and LRP1B in Figure 5. High and low expressions of the four genes were summarized (Figure 5). High expressions of RNF213, KMT2D and CSMD3 were observed in lung cancer tissues, and low expressions of these genes were observed in benign disease tissues. There were significant statistical differences of these genes between the two groups (P<0.05), especially RNF213 (P<0.005). Low expression of LRP1B was observed in 26

lung cancer tissues and in 14 benign disease tissues. One of the lung cancer samples was high expression. This result was not significantly different (P>0.05).

Validation set results

There were 28 lung nodule patients enrolled the validation set without a definite diagnosis through the chest CT. The largest diameter of the lesion was all less than 3 cm in CT imaging. Malignant or benign nodules could not be confirmed in CT imaging and biomarkers. Blood samples were targeted sequenced in the same method. The number of the test genes RNF213, KMT2D, CSMD3 and LRP1B detected in lung cancer samples was 5, 5, 3 and 2 respectively. Twenty samples were confirmed to be lung cancer and 8 samples were benign nodules with postoperative pathology. RNF213 gene mutated in 25% lung cancer patients. KMT2D, CSMD3 and LRP1B genes mutated in 15%, 10% and 10% lung cancer patients, but KMT2D and CSMD3 genes were detected in 25% and 12.5% benign diseases (Figure 6).

An analysis of GGO and GGN in all samples

We put all sequenced data of the 70 samples in an analysis to determine the veracity of the result. There were 55 patients diagnosed GGO or GGN in chest CT including 36 early stage lung cancer and 19 benign disease. We detected RNF213 gene mutation in 10 (10/36, 27.8%) lung cancer samples and no samples in benign disease (P<0.05). All of these somatic mutations were missense mutation. The specificity of RNF213 gene mutation was 100% in diagnosis of GGO and GGN, and its sensibility was 27.8%.

Discussion

Early diagnosis and treatment are effective means to improve the survival rate. Small lesion of lung can be found in chest CT which is the most common and valid examination in use of diagnosis or screening in lung cancer[25, 26]. Recently, early detection or screening with low dose computed tomography (LDCT) was shown to improve survival and reduce lung cancer specific mortality by the National Lung screening Trial (NLST) and other studies[27,28].

Some lesions in CT are easy to diagnose lung cancer, some lesions are difficult to be identified for a lung cancer or benign disease especially when the lung nodule is small or imaging features are not typical[29]. For example, the ground-glass opacity (GGO) or ground-glass nodule (GGN) is usually adenocarcinoma in situ (AIS), minimally invasive adenocarcinoma (MIA) or atypical adenomatous hyperplasia (AAH)[30]. Higher false positive and false negative rate by using traditional biomarkers like carcino-embryonic antigen (CEA), neuro-specific enolase (NSE) and cytokeratin 19 (CYFRA-211) make the diagnosis more

difficult[31,32]. These biomarkers are useless in early diagnosis of lung cancer. Aspiration biopsy may be needed to confirm the nodule is a malignancy or benign disease. But, bleeding, pneumothorax, pain and possible diversion restrict its use in early diagnosis.

An ideal diagnostic method should be simple and convenient, safe and efficient. Using liquid biopsies to detect circulating biomarkers such as circulating tumor cell (CTC), Circulating tumor DNA (ctDNA) and exosome may offer a relatively simple method to analyze early stage tumor[33]. Detecting ctDNA in peripheral blood was used more commonly, while CTC was less commonly detected in early stage tumor.

Circulating cell-free DNA, a fragment of DNA which is released through cell apoptosis, widely exists in extracellular fluid such as blood, cerebrospinal fluid, urine and saliva[4,5]. The cfDNA of healthy people comes mainly from metabolism and cell apoptosis including bone marrow cells, lymphocytes, and normal tissue cells[34]. For patients with tumors, the cfDNA fragments of tumor cells known as circulating tumor DNA(ctDNA) also were released to peripheral blood through apoptosis and necrosis of tumor cells[6,7]. Plasma ctDNA, which is a fragment about 150-200 bp[35] containing genetic information about the tumor, is of great significance to the diagnosis, treatment and monitoring of the disease.

Circulating tumor DNA was used for monitoring therapeutic effect and prognostic prediction in treatment of malignancy because of ctDNA level in advanced stage tumor[36].The levels of detected ctDNA increase correlate with the malignant progression[37,13]. Low level of ctDNA in early stage tumor makes the detection difficult. Early diagnosis can provide tremendous benefits for the treatment of patients with malignant tumors[13,38]. With the development of sequencing technology, low level ctDNA could be detected in blood more easily and accurately. More and more studies were applied this technology to investigate early stage tumors.

In the present study, we investigated ctDNA in early stage lung cancer and comparable benign disease diagnosed in chest CT. First, cfDNA could be detected in all lung benign disease samples as being reported in other solid tumors previously[39,40,41]. However, these studies did not consider the stage of malignancy compared to benign disease. The levels of cfDNA in malignancy related with the tumor burden such as tumor size, T stage and TNM stage[42]. Our data indicated that the level of cfDNA in early stage lung cancer was not significantly different with benign disease (0.53±0.66 ng/μl and 0.54±0.29 ng/μl, respectively; $P>0.05$). This result may be related to low tumor burden in early stage lung cancer, indicating that early stage lung cancers release low level cfDNA into blood stream similarly with lung benign disease. Cell apoptosis and necrosis from benign tumor or disease also cause cfDNA increasing.

Elevated cfDNA concentrations alone did not fully distinguish between lung cancer and benign disease. In our study, targeted NGS was implemented to detect ctDNA in these DNA samples. The panel used for NGS covered all known mutated genes in malignant tumor to investigate mutations in early stage lung cancer. We found a number of mutations were not related to gender, age, smoking history, tumor size, stage or pathology in two groups. Some genes mutated more frequently in lung cancer and others in benign disease. In training set, RNF213, KMT2D, CSMD3 and LRP1B genes mutated more frequently in

early stage lung cancer than in benign diseases. There were 25.9% lung cancer patients showed RNF213 gene mutation and no one in benign disease patients. RNF213 gene has a high specificity in lung cancer and benign disease.

In order to make clear the protein expression of these four genes in tissues, we conducted immunohistochemistry of lung cancer tissues. RNF213, KMT2D and CSMD3 genes showed a higher expression than in benign disease samples especially RNF213. This differential expression in two groups may be due to the change in amino acid caused by genetic changes.

Finally, a verification experiment was carried out to study the diagnosis effect of these four genes. RNF213 gene mutated in twenty five percent of lung cancer patients and not mutated in benign diseases, but KMT2D, CSMD3 and LRP1B mutated less in both two groups. The same high specificity of RNF213 gene mutation was showed in the validation set, although the difference was not statistically significant, probably due to a small number of samples. In all 70 samples including training set and validation set, RNF213 gene mutation was significantly different compared with lung benign diseases. In addition, we investigated RNF213 in all GGO and GGN patients of the study. There were 27.8% lung cancer samples showed RNF213 mutation and no samples in benign disease ($P<0.05$) with the similar high specificity. A larger number of randomized controlled samples need to be studied to further confirm these results.

RNF213 gene, known as ring finger protein 213, encodes a protein containing a RING finger domain[43]. It was found in some malignant tumors such as ovarian cancer, gastric cancer and liver cancer[44,45,46], yet there were a few studies about RNF213 gene mutation in malignant tumors. RNF213 has been reported that it may be a tumor suppressor in malignancy[47]. We first found RNF213 gene mutation in ctDNA of early stage lung cancer, and it was significantly statistical difference compared with lung benign disease. The missense mutation of RNF213 changed the amino acid thus affecting the protein function. This gene mutation resulted in a loss of its function of tumor suppressor and promoted tumor development and progression in lung cancer. The novel mechanisms need to be lucubrated in the future.

Conclusions

In conclusions, the concentration of cfDNA cannot be a good biomarker in diagnosis of early stage lung cancer and lung benign diseases. RNF213 gene mutation of ctDNA may be used for molecular diagnosis of lung malignant and benign nodules through targeted NGS. The effect of KMT2D, CSMD3 and LRP1B should be further confirmed in more samples.

To the best of our knowledge, this is the first tie to report that RNF213 gene mutation in ctDNA was detected by targeted NGS in early stage lung cancer. It has high specificity of 100% and sensibility of about 26% in diagnosis or screening of lung cancer. A larger-scale randomized controlled trial is needed to verify this finding in the future. In addition, the underlying mechanisms of these genes causing

recurrence and development of lung cancer need a further study. The results of our study would be useful in the diagnosis and treatment of early stage lung cancer.

Abbreviations

NGS: Next-generation sequence; cfDNA: cell free DNA; ctDNA: circulating tumor DNA; CEA: carcino-embryonic antigen; NSE: neuro-specific enolase; CYFRA-211: cytokeratin 19; SNV: single nucleotide variation; InDel: insertion-deletion; LDCT: low dose computed tomography

Declarations

Acknowledgements

The authors would like to thank Wenjuan Wang (The Second Hospital of Shandong University) for advice during preparation of this manuscript.

Authors' contributions

NJ analyzed the data, consulted literature, and wrote the manuscript. SK Z managed the samples. YL, HB S, WH Z, YT H and YP Z collected blood samples. CK Z and CJ Z carried out the experiment. PC L, CL P, QF S and WQ Z collected and interpreted data. YS W advised for experiment modification. XG Z designed and managed the study, and reviewed the manuscript. All authors read and approved the final manuscript.

Funding

This study was supported in part by the Key Research and Development Program of Shandong Province (Grant No. 2016ZDJS07A15, 2017G006028, 2017GSF22107).

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

This study was approved by the Ethics Committee of the Second Hospital of Shandong University (KYLL-2016(LW)-0025). Written informed consent was obtained from all patients or their guardians.

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Author details

[1] Department of Thoracic Surgery, The Second Hospital of Shandong University, Jinan, Shandong 250033, P.R.China

[2] Key Laboratory of Chest Cancer, Shandong University, Jinan, Shandong 250033, P.R.China

[3] Cheeloo College of Medicine, Shandong University, Jinan, Shandong 250012, P.R.China

[4] The 960th Hospital of People's Liberation Army of China, Jinan, Shandong 250031, P.R.China

[5] Shandong Provincial Chest Hospital, Jinan, Shandong 250013, P.R.China

[6] Department of Clinical Laboratory, The Second Hospital of Shandong University, Jinan, Shandong 250033, P.R.China

[7] Pathology Department, The Second Hospital of Shandong University, Jinan, Shandong 250033, P.R.China

[8] Department of Thoracic Surgery, Shandong Provincial Hospital, Jinan, Shandong 250021, P.R.China

References

1. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. CA Cancer J Clin. 2013;63(1):11-30.
2. Richards TB, Henley SJ, Puckett MC, Weir HK, Huang B, Tucker TC, Allemani C. Lung cancer survival in the United States by race and stage (2001-2009): Findings from the CONCORD-2 study. Cancer. 2017 Dec 15;123 Suppl 24:5079-5099.

3. Torre LA, Siegel RL, Jemal A. Lung Cancer Statistics. Adv Exp Med Biol. 2016;893:1-19.
4. Wang Y, Springer S, Mulvey CL, Silliman N, Schaefer J, et al. Detection of somatic mutations and HPV in the saliva and plasma of patients with head and neck squamous cell carcinomas. Sci Transl Med. 2015;7(293):293ra104.
5. De Mattos-Arruda L, Mayor R, Ng CKY, Weigelt B, Martinez-Ricarte F, et al. Cerebrospinal fluid-derived circulating tumour DNA better represents the genomic alterations of brain tumours than plasma. Nat Commun. 2015;6:8839.
6. Jahr S, Hentze H, Englisch S, Hardt D, Fackelmayer FO, et al. DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. Cancer Res. 2001;61(4):1659-65.
7. El Messaoudi S, Rolet F, Mouliere F, Thierry A R. Circulating cell free DNA: preanalytical considerations. Clin Chim Acta. 2013;424:222-30.
8. Pi C, Zhang MF, Peng XX, Zhang YC, Xu CR, Zhou Q. Liquid biopsy in non-small cell lung cancer: a key role in the future of personalized medicine? Expert Rev Mol Diagn. 2017 Dec;17(12):1089-1096.
9. Schwarzenbach H, Hoon DS, Pantel K. Cell-free nucleic acids as biomarkers in cancer patients. Nat Rev Cancer. 2011 Jun;11(6):426-37.
10. Murtaza M, Dawson SJ, Tsui DW, Gale D, Forshew T, et al. Non-invasive analysis of acquired resistance to cancer therapy by sequencingof plasma DNA. Nature. 2013;497(7447):108-12.
11. Gevensleben H, Garcia-Murillas I, Graeser MK, Schiavon G, Osin P, et al. Noninvasive detection of HER2 amplification with plasma DNA digital PCR. Clin Cancer Res. 2013;19(12):3276-84.
12. Diehl F, Li M, Dressman D, He Y, Shen D, et al. Detection and quantification of mutations in the plasma of patients with colorectal tumors. Proc Natl Acad Sci U S A. 2005;102(45):16368-73.
13. Bettegowda C, Sausen M, Leary RJ, Kinde I, Wang Y, et al. Detection of circulating tumor DNA in early- and late-stage human malignancies. Sci Transl Med. 2014;6(224):224ra24.
14. Newman AM, Bratman SV, To J, Wynne JF, Eclov NC, et al. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. Nat Med. 2014;20(5):548-54.
15. Chen KZ, Lou F, Yang F, Zhang JB, Ye H , et al. Circulating Tumor DNA Detection in Early-Stage Non-Small Cell Lung Cancer Patients by Targeted Sequencing. Sci Rep. 2016;6:31985.
16. Guo N, Lou F, Ma Y, Li J, Yang B, et al. Circulating tumor DNA detection in lung cancer patients before and after surgery. Sci Rep. 2016;6:33519.
17. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010 Mar 1;26(5):589-595.
18. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009 Aug 15;25(16):2078-2079.
19. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol. 2013 Mar;31(3):213-219

20. Saunders CT, Wong WS, Swamy S, Becq J, Murray LJ, Cheetham RK. Strelka: accurate somatic small-variant calling from sequenced tumor–normal sample pairs. Bioinformatics. 2012 Jul 15;28(14):1811-1817.
21. 1000 Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012 Nov 1;491(7422):56-65.
22. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016 Aug 18;536(7616):285-291.
23. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010 Sep;38(16):e164.
24. Kawai H, Ishii A, Washiya K, Konno T, Kon H, et al. Estrogen receptor alpha and beta are prognostic factors in non-small cell lung cancer. Clin Cancer Res. 2005;11(14):5084-9.
25. Latimer KM, Mott TF. Lung cancer: diagnosis, treatment principles, and screening. Am Fam Physician. 2015 Feb 15;91(4):250-6.
26. Tanoue LT. Lung cancer screening. Curr Opin Pulm Med. 2016 Jul;22(4):327-35.
27. National Lung Screening Trial Research Team, Aberle DR, Berg CD, Black WC, Church TR, Fagerstrom RM, et al. The National Lung Screening Trial: overview and study design. Radiology. 2011;258(1):243-53.
28. Pastorino U, Silva M, Sestini S, Sabia F, Boeri M, et al. Prolonged Lung Cancer Screening Reduced 10-year Mortality in the MILD Trial. Ann Oncol. 2019.pii:mdz117.
29. Lee CT. What do we know about ground-glass opacity nodules in the lung? Transl Lung Cancer Res. 2015 Oct;4(5):656-9.
30. Travis WD, Brambilla E, Noguchi M, Nicholson AG, Geisinger KR, Yatabe Y, et al. . International association for the study of lung cancer/american thoracic society/european respiratory society international multidisciplinary classification of lung adenocarcinoma. J Thorac Oncol. 2011 Feb;6(2):244-85.
31. Wang B, He YJ, Tian YX, Yang RN, Zhu YR, Qiu H. Clinical utility of haptoglobin in combination with CEA, NSE and CYFRA21-1 for diagnosis of lung cancer. Asian Pac J Cancer Prev. 2014;15(22):9611-4.
32. Wang P, Piao Y, Zhang X, Li W, Hao X. The concentration of CYFRA 21-1, NSE and CEA in cerebro-spinal fluid can be useful indicators for diagnosis of meningeal carcinomatosis of lung cancer. Cancer Biomark. 2013;13(2):123-30.
33. Pantel K, Alix-Panabières C. Real-time liquid biopsy in cancer patients: fact or fiction? Cancer Res. 2013;73(21):6384-8.
34. Atamaniuk J, Kopecky C, Skoupy S, Säemann MD, Weichhart T. Apoptotic cell-free DNA promotes inflammation in haemodialysis patients. Nephrol Dial Transplant. 2012;27(3):902-5.

35. Pös O, Biró O, Szemes T, Nagy B. Circulating cell-free nucleic acids: characteristics and applications. Eur J Hum Genet, 2018;26(7):937-945.
36. Stewart CM, Tsui DWY. Circulating cell-free DNA for non-invasive cancer management. Cancer Genet. 2018;228-229:169-179.
37. Rumiao E, Boldrin E, Malacrida S, Realdon S, Fassan M, et al. Detection of genetic alterations in cfDNA as a possible strategy to monitor the neoplastic progression of Barrett's esophagus. Transl Res. 2017;190:16-24.e1.
38. Tie J, Kinde I, Wang Y, Wong HL, Roebert J, et al. Circulating tumor DNA as an early marker of therapeutic response in patients with metastatic colorectal cancer. Ann Oncol. 2015;26(8):1715-22.
39. Shapiro B, Chakrabarty M, Cohn EM, Leon SA. Determination of circulating DNA levels in patients with benign or malignant gastrointestinal disease. Cancer. 1983;51(11):2116-20.
40. Qiu YW, Shen XJ, Jin CJ, Cao XJ, Ju SQ. Value of the concentration and integrity of serum cell-free DNA for the clinical diagnosis of esophageal carcinoma. Zhonghua Zhong Liu Za Zhi.2018;40(12):905-910.
41. Ponti G, Maccaferri M, Micali S, Manfredini M, Milandri R, et al. Seminal Cell Free DNA Concentration Levels Discriminate Between Prostate Cancer and Benign Prostatic Hyperplasia. Anticancer Res. 2018;38(9):5121-5125.
42. Kim K, Shin DG, Park MK, Baik SH, Kim TH, et al. Circulating cell-free DNA as a promising biomarker in patients with gastric cancer: diagnostic validity and significant reduction of cfDNA after surgical resection. Ann Surg Treat Res. 2014;86(3):136-42.
43. Liu W, Morito D, Takashima S, Mineharu Y, Kobayashi H, Hitomi T, et al. Identification of RNF213 as a susceptibility gene for moyamoya disease and its possible role in vascular development. PLoS One. 2011;6(7):e22542.
44. Er TK, Su YF, Wu CC, Chen CC, Wang J, Hsieh TH, et al. Targeted next-generation sequencing for molecular diagnosis of endometriosis-associated ovarian cancer. J Mol Med (Berl). 2016;94(7):835-47.
45. Ge S, Li B, Li Y, Li Z, Liu Z, Chen Z, et al. Genomic alterations in advanced gastric cancer endoscopic biopsy samples using targeted next-generation sequencing. Am J Cancer Res. 2017;7(7):1540-1553.
46. Li X, Xu W, Kang W, Wong SH, Wang M, et al. Genomic analysis of liver cancer unveils novel driver genes and distinct prognostic features. Theranostics. 2018;8(6):1740-1751.
47. Banh RS, Iorio C, Marcotte R, Xu Y, Cojocari D, Rahman AA, et al. PTP1B controls non-mitochondrial oxygen consumption by regulating RNF213 to promote tumour survival during hypoxia. Nat Cell Biol. 2016;18(7):803-813.

Tables

Table 1: Patients’ general characteristics in training set

	Lung cancer	Benign disease		P value
	No. (%)	No. (%)		
Number	27 (64.3%)	15 (35.7%)		
Gender				>0.05
Male	18 (42.9%)	10 (23.8%)		
Female	9 (21.4%)	5 (11.9%)		
Age	44-73	33-72		<0.05
Smoking				>0.05
Smoker	13 (31.0%)	6 (14.3%)		
Nonsmoker	14 (33.3%)	9 (21.4%)		
Tumor Size				>0.05
≤3cm	18 (42.9%)	11 (26.2%)		
3cm□T*≤5cm	9 (21.4%)	4 (9.5%)		
Pathology				
Ad	18 (42.9%)	tuberculosis	3 (7.1%)	
SC	7 (16.7%)	inflammation	9 (21.4%)	
Others	2 (4.8%)	hamartoma	2 (4.8%)	
		SH	1 (2.4%)	
Stage				
IA	5 (11.9%)			
IB	14 (33.3%)			
IIA	6 (14.3%)			
IIB	2 (4.8%)			

*T= Tumor□Ad=Adenocarcinoma□SC=Squamous carcinoma□SH=sclerosing hemangioma

Others=Big cell carcinoma, sarcomatoid carcinoma

Table 2: Positive biomarkers of two group in training set and validation set

Biomarker	Training set			Validation set		
	LC	BD	P value	LC	BD	P value
CEA	3/17	0/12	>0.05	1/18	0/8	>0.05
NSE	8/16	6/12	>0.05	4/18	1/8	>0.05
CYFRA-211	2/13	0/12	>0.05	3/18	1/8	>0.05

LC=Lung cancer BD=benign disease

Table 3: Mutation sites of significant genes in lung cancer

Gene	Chromosome	Exon	cDNA	Amino acid	Type	Num.
RNF213	17	26	c.G5960A	p.G1987E	Missense	1
	17	26	c.C5171T	p.A1724V	Missense	2
	17	26	c.C5155T	p.Q1719X	Nonsense	1
	17	59	c.C14226A	p.S4742R	Missense	1
	17	29	c.A8252T	p.N2751I	Missense	1
	17	17	c.A3101T	p.K1034M	Missense	1
	17	24	c.G4615A	p.A1539T	Missense	1
KMT2D	12	34	c.C8972T	p.P2991L	Missense	1
	12	34	c.C8495T	p.A2832V	Missense	2
	12	16	c.C4474T	p.Q1492X	Nonsense	1
	12	4	c.G487A	p.A163T	Missense	1
	12	41	c.G13711A	p.A4571T	Missense	1
	12	11	c.C3647T	p.A1216V	Missense	1
	12	38	c.G10607A	p.R3536H	Missense	1
CSMD3	8	54	c.C8300T	p.P2767L	Missense	1
	8	20	c.A3058T	p.T1020S	Missense	3
	8	14	c.C2050T	p.L684F	Missense	1
	8	10	c.G1397C	p.S466T	Missense	1
	8	59	c.G9323T	p.G3108V	Missense	1
LRP1B	2	51	c.G8185A	p.A2729T	Missense	1
	2	83	c.A12658T	p.T4220S	Missense	1
	2	2	c.A125T	p.H42L	Missense	3
	2	16	c.G2555A	p.C852Y	Missense	1
	2	89	c.A13511T	p.H4504L	Missense	1
	2	8	c.A1141T	p.N381Y	Missense	1

Figures

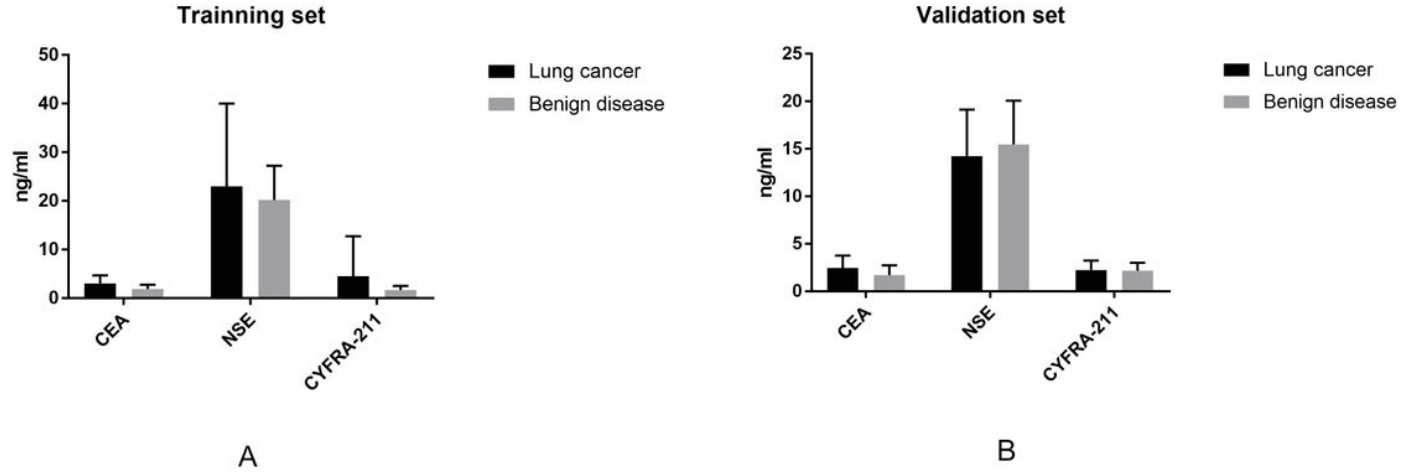

Figure 1

The concentration of biomarker including CEA, NSE and CYFRA-211. In training set (A) and validation set (B), there were no significant statistical differences between lung cancer and benign disease group respectively in training set and validation set (P>0.05).

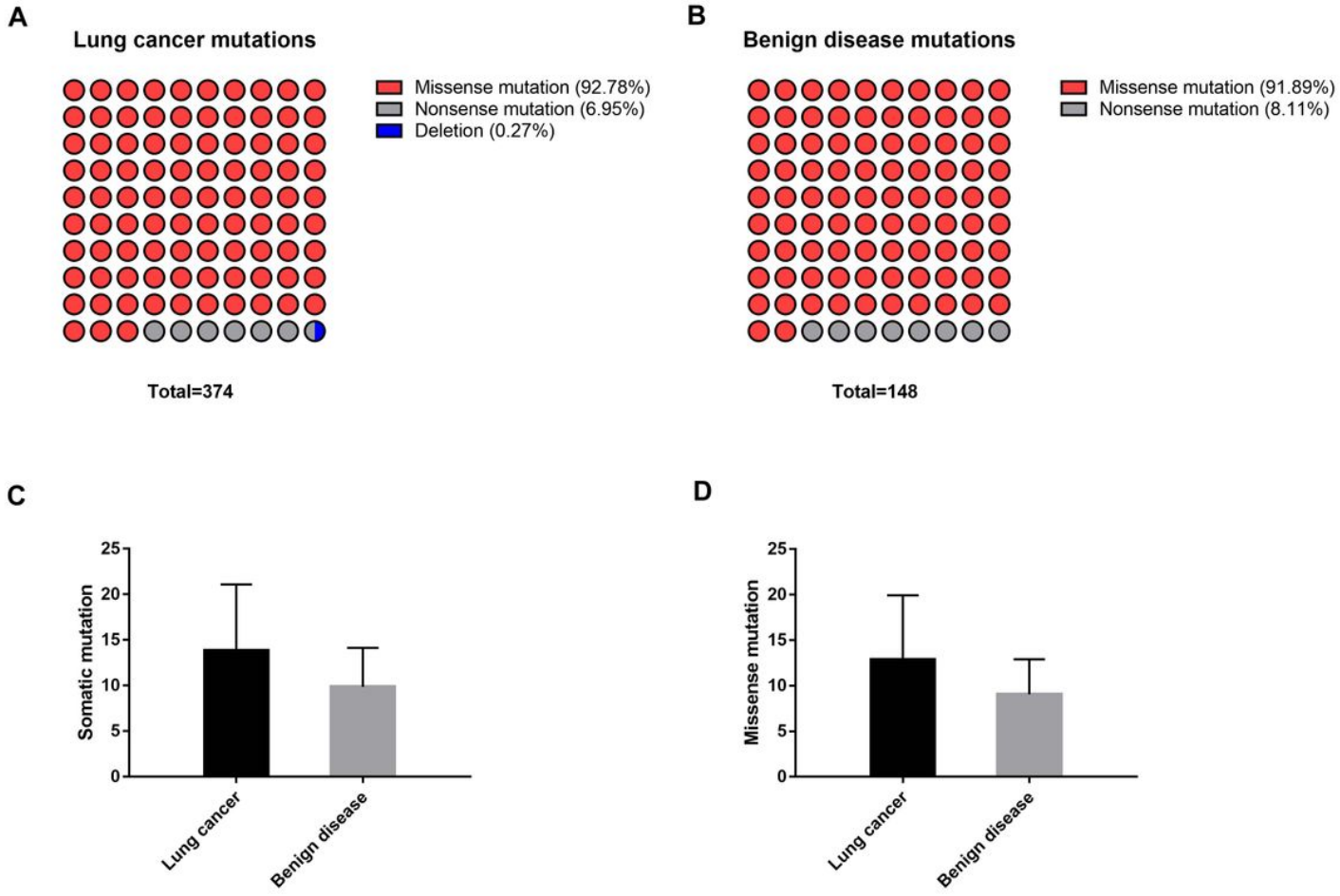

Figure 2

The mutations in training set. (A) In lung cancer group, 92.78 percent mutations were missense mutations, 6.95 percent mutations were nonsense mutations and 0.27 percent mutations were deletions. (B) In lung benign disease group, 91.89 percent mutations were missense mutations and 8.11 percent mutations were nonsense mutations. The number of somatic mutations (C) and missense mutations (D) in two groups, and there were no significant statistical differences between lung cancer and benign disease (P>0.05).

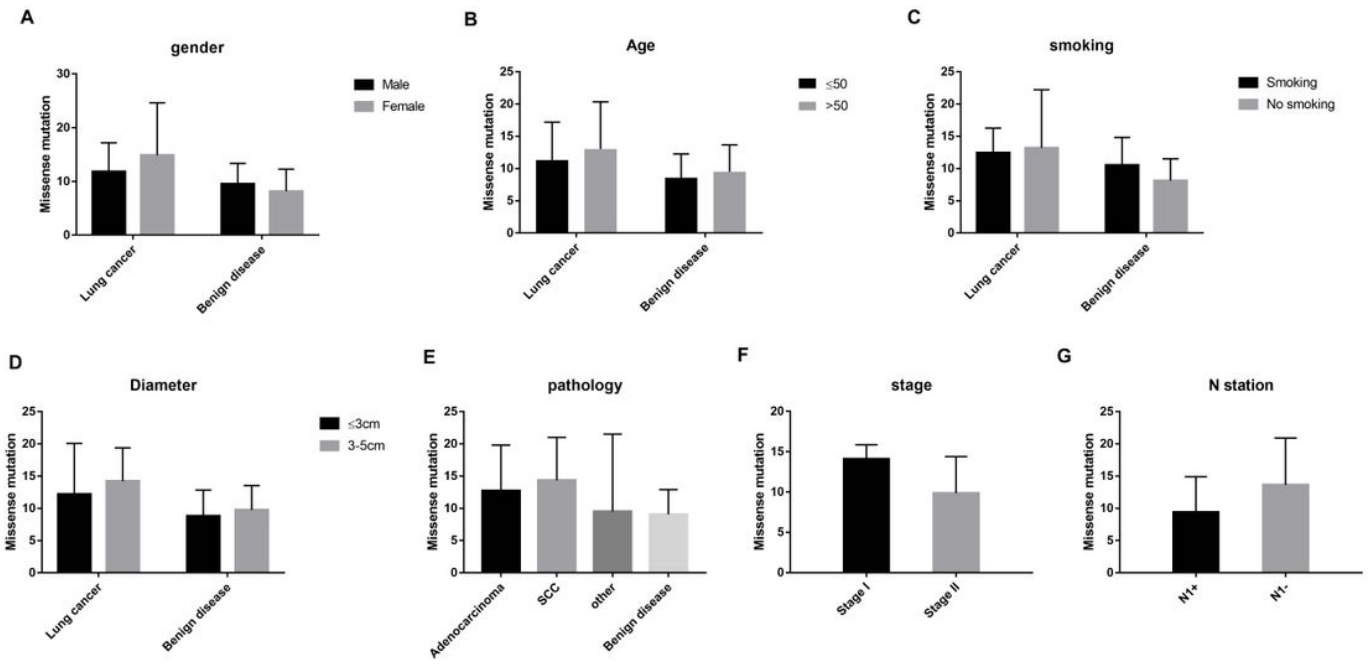

Figure 3

The number of missense mutations in two groups. (A) In lung cancer group, mutations of male and female respectively was 11.83±5.36 and 14.89±9.74. In lung benign disease group, mutations of male and female respectively was 9.5±3.86 and 8.20±4.09. (B) In lung cancer group, mutations of age less than 50 years respectively was 11.33±5.86 and 8.57±3.69. (C) Mutations in lung cancer group, smoking patients was 12.42±3.85 and no smoking patients was 13.20±9.01. Mutations in benign disease, smoking patients was 10.50±4.32 and no smoking patients was 8.11±3.41. (D) Mutations in lung cancer less than 3 cm and 3-5 cm group respectively was 12.17±7.91 and 14.22±5.17. Mutations in benign disease less than 3 cm and 3-5 cm group respectively was 8.82±4.02 and 9.75±3.77. (E) Mutations in adenocarcinoma, SCC, others and benign disease was 12.74±7.07, 14.33±6.68, 9.50±12.02 and 9.07±3.84. (F) The number of missense mutations in stage I and stage II lung cancer was 14.10±7.67 and 9.87±4.52. (G) The number of missense mutations in N1 station lymph node positive and negative respectively was 9.40±5.50 and 13.64±7.27. There were no statistically significant difference in all the comparisons (P>0.05).

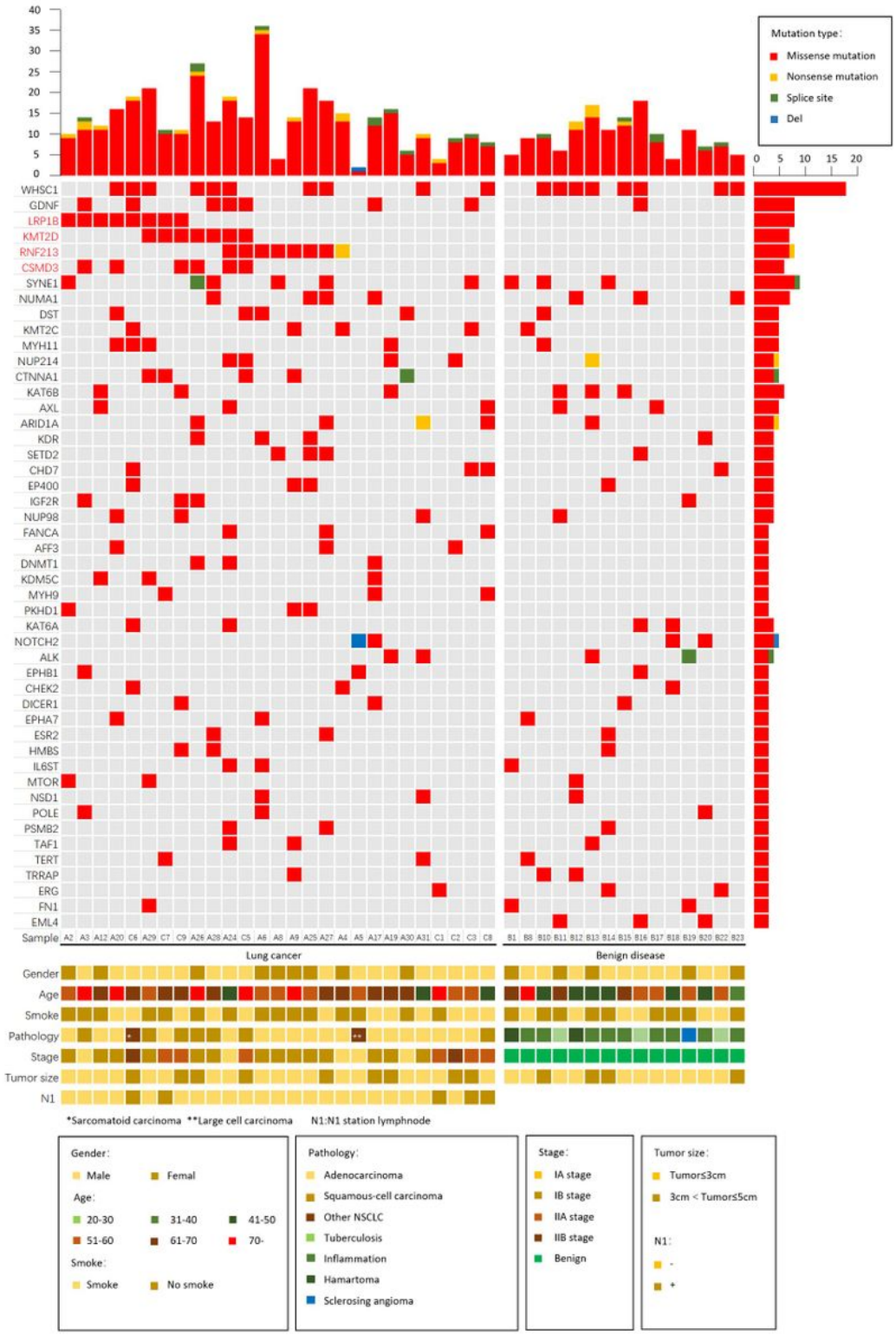

Figure 4

Heat map of somatic SNVs in lung cancer and benign disease group of the training set. The most frequently mutated gene was WHSC1. There were 10 and 8 samples mutated respectively in two groups (P>0.05). RNF213, KMT2D, CSMD3 and LRP1B were more frequently mutated in lung cancer than benign disease (P<0.05). The genes mutated less than three were not listed in the heat map (All data was shown in supplementary).

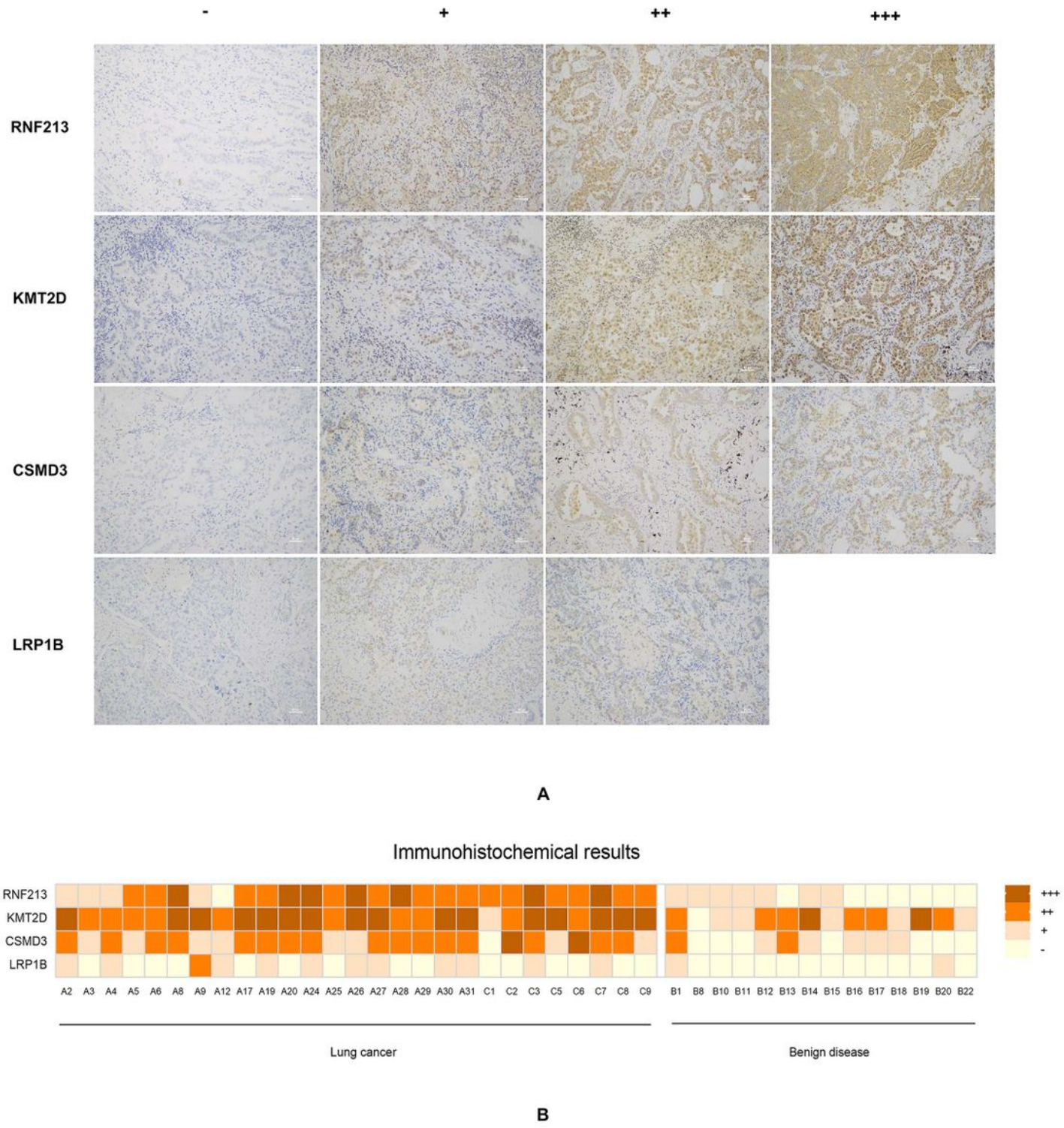

Figure 5

The results of immunohistochemistry. (A) Representative illustrations of expressions of RNF213, KMT2D, CSMD3 and LRP1B, in lung cancer and benign disease. “-” and “+” represent low expression of RNF213, KMT2D, CSMD3 and LRP1B (IHC, ×200). “++” and “+++” represent high expression of this four genes (IHC, ×200). (B) The summary of RNF213, KMT2D, CSMD3 and LRP1B expressions in 27 lung cancer and 14 benign disease tissue

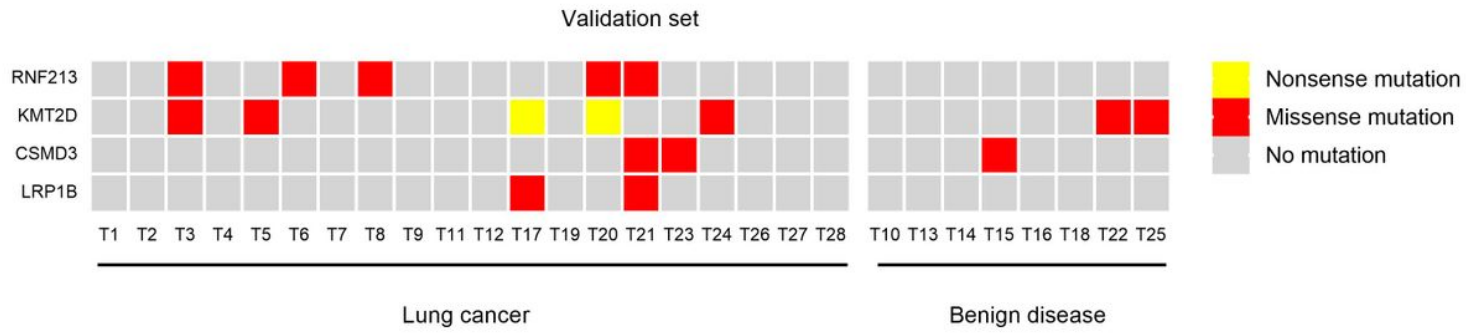

Figure 6

Mutations of validation set. RNF213 gene mutations were detected in five lung cancer samples and no one in benign control group. RNF213 gene mutation has high specificity of 100% and sensibility of about 25%. It is consistent with the data in the training set.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FigureS1.tif
- TableS2.xlsx
- TableS1.xls