ART & TECHNOLOGY PUBLISHING PTE. LTD.

Cell Biology Research
2024 Volume 5, Issue 1
ISSN: 2529-7627(Online)
ISSN: 2529-7643(Print)

Lung Adenocarcinoma Gene Mutation in Koreans Detection Using Next Generation Sequence Analysis Technique and Analysis of Concordance with Existing Genetic Test Methods

Jae Ha Baek, Kyu Bong Cho*

Department of Biomedical Laboratory Science, Shinhan University, Uijeongbu 11644, Republic of Korea

***Corresponding author:** Kyu Bong Cho, kbcho@shinhan.ac.kr

Abstract

Lung adenocarcinoma accounts for about 40% of all lung cancers. With the recent development of gene profiling technology, studies on mutations in oncogenes and tumor suppressor genes, which are important for the development and growth of tumors, have been actively conducted. Companion diagnosis using next-generation sequencing helps improve survival with targeted therapy. In this study, formalin-fixed paraffin-embedded tissues of non-small cell lung cancer patients were subjected to hematoxylin and eosin staining to detect genetic mutations that induce lung adenocarcinoma in Koreans. Immunohistochemical staining was also performed to accurately classify lung adenocarcinoma tissues. Based on the results, next-generation sequencing was applied to analyze the types and patterns of genetic mutations, and the association with smoking was established as the most representative cause of lung cancer. The results of next-generation sequencing analysis confirmed the single nucleotide variations, copy number variations, and gene rearrangements. To validate the reliability of next-generation sequencing, we additionally performed the existing genetic testing methods (polymerase chain reaction-epidermal growth factor receptor, immunohistochemistry-anaplastic lymphoma kinase [D5F3], and fluorescence *in situ* hybridization-receptor tyrosine kinase 1 tests) to confirm the concordance rates with the next-generation sequencing test results. This study demonstrates that next-generation sequencing of lung adenocarcinoma patients simultaneously identifies mutation.

Keywords

Adenocarcinoma
Lung cancer
Next-generation sequencing
Non-small cell lung cancer
Small cell lung cancer

1. Introduction

According to data released by the Centralized Cancer Registry in 2021, in South Korea in 2021, there were a total of 28,628 cases of lung cancer, accounting for 11.8% of all cancer cases and ranking second. The gender distribution was 2.1:1, with more cases occurring in men, totaling 20,331 cases, making it the most common cancer in men, while women had 9,629 cases, ranking fifth among female cancers. Notably, lung cancer cases have been increasing by an average of 3.2% per year among women, while the increase in men has been gradually decreasing [1].

Lung cancer has been primarily attributed to smoking, as cigarettes contain over 7,000 harmful substances, with more than 60 known carcinogens. Smokers face an 11-fold higher risk of developing lung cancer compared to non-smokers, and the risk increases with the number of cigarettes smoked, the duration of smoking, and the age at which one starts smoking. In South Korea, 69.8% of lung cancer cases were directly linked to smoking, and in countries with high smoking rates like the United States and the European Union, over 90% of male lung cancer cases and 74% to 80% of female lung cancer cases were attributed to smoking [2-5].

Lung cancer is broadly categorized into small cell carcinoma and non-small cell carcinoma. Data from international cancer research institutions indicate that non-small cell carcinoma accounts for approximately 70% to 80% of lung cancer cases, with variations in the gender distribution within subtypes. For instance, lung squamous cell carcinoma is more common in men, while lung adenocarcinoma is more common in women [6]. Lung cancer has a very low survival rate, with a 5-year survival rate of 18.6%, significantly lower than other major cancer types such as colorectal cancer (64.5%), breast cancer (89.6%), and prostate cancer (98.2%). In South Korea, the 5-year survival rate for both men and women is around 20%, highlighting the urgent need for improved diagnostic and treatment methods [7].

However, recent advancements in genetic profiling technology have led to significant improvements in the understanding of the genetic mutations that drive tumor development and growth. Specific genetic mutations associated with lung cancer have been discovered, which have had a substantial impact on survival rates [8]. Unlike the previously used polymerase chain reaction (PCR), fluorescence in situ hybridization (FISH), and immunohistochemistry (IHC) testing methods, new molecular testing techniques, such as next-generation sequencing (NGS), can identify mutations in tumor genes and tumor suppressor genes. NGS can simultaneously identify point mutations, insertions, deletions, fusions, translocations, and copy number variations (CNVs) by sequencing a target deoxyribonucleic acid (DNA) locus, the entire exome, or the entire genome in a single test. Confirmation of these genetic mutations through NGS can lead to personalized targeted therapies that significantly improve survival rates. In particular, lung adenocarcinoma can benefit from personalized medicine through companion diagnostics using NGS [9, 10]. Mutations in the epidermal growth factor receptor (EGFR) gene, a driver oncogene in lung adenocarcinoma, have shown favorable responses to EGFR tyrosine kinase inhibitor (TKI) treatment, resulting in higher survival rates [11]. It has also been reported that anaplastic lymphoma kinase (ALK) and receptor tyrosine kinase 1 (ROS1) also exhibit better survival with ALK TKI treatment [11]. Thus, the response rate to appropriately targeted therapy in lung cancer patients identified by molecular testing techniques such as NGS is high (60% to 70%) compared to the response rate in the untargeted patient population (20% to 30%) [12].

Therefore, it is essential to perform genetic testing to analyze each patient's genetic mutation pattern and apply it to targeted therapy to improve their survival rates. Recently, NGS has become a valuable tool for rapidly analyzing genetic sequences using formalin-fixed paraffin-embedded (FFPE) tissues, allowing for efficient analysis of the genetic sequences of many target genes and facilitating personalized treatment for patients [13].

In this study, 83 patients diagnosed with lung adenocarcinoma between 2018 and 2020 underwent NGS analysis to detect single nucleotide variations (SNVs), CNVs, and structural variations (SVs) in two tumor tissues. Conventional genetic mutation detection methods such as PCR, FISH, and IHC staining were also performed. The study evaluated the utility of NGS for the detection of genetic mutations in lung cancer patients and analyzed the concordance between NGS results and the results of conventional genetic mutation detection methods.

2. Materials and methods

2.1. Collection of clinical specimen and patient data

We utilized FFPE samples prepared from surgical resection or biopsy tissues of 83 patients diagnosed with lung cancer at Hanyang University Hospital between 2018 and 2020. All patients who had signed informed consent forms for various diagnostic and genetic tests were included in this study. Patient data for research purposes were collected from medical records while ensuring that all personal information was encrypted to prevent identification. The study received ethical approval from the Hanyang University Hospital Institutional Review Board (IRB approval number: HYUH 2022-03-030-001).

2.2. Hematoxylin and eosin staining

Hematoxylin and eosin (H&E) staining was performed to differentiate lung cancer based on tissue differentiation patterns. FFPE tissue sections, cut into 4 μm thickness, were mounted on slides using a paraffin dispenser and melted at 62°C for 30 min. Paraffin was then removed with xylene, followed by a series of alcohol washes (100%, 95%, 80%, 50%), and rinsed with water. Subsequently, the tissue sections were stained with hematoxylin (DAKO, Santa Clara, CA, USA) for 5 minutes to visualize cell nuclei. After a water rinse, a bluing buffer (DAKO) was used for counterstaining, and eosin (DAKO) was applied to stain the cytoplasm. The tissue sections were dehydrated sequentially in 50%, 80%, 95%, and 100% ethanol, cleared in xylene, and finally mounted for microscopic examination. Based on the differentiation patterns and morphological criteria of lung cancer tissue, differentiation was classified as well-differentiated, moderately differentiated, or poorly differentiated.

2.3. Immunohistochemistry staining

To accurately classify lung cancer, IHC staining was performed. FFPE sections, cut to a thickness of 4 μm, were stained using the Bond max III (Leica Biosystems, Buffalo Grove, IL) and Benchmark automatic immunostaining device (Ventana Medical System, Tucson, AZ, USA). The FFPE sections were attached to slides and melted at 62°C for 30 min to remove excess paraffin, followed by deparaffinization in xylene. The sections were then dehydrated sequentially in 100%, 95%, and 75% graded alcohols and rinsed in distilled water. Subsequently, they were subjected to a 15-minute reaction in 3% hydrogen peroxide solution to inhibit endogenous peroxidase activity.

A blocking antibody (100 μL) was applied to the tissue sections and allowed to react at room temperature for 30 min. The primary antibody was then diluted and reacted at room temperature. After washing with pH 7.6 Tris-HCl buffer, 100 μL of biotinylated link antibody (LSAB kit; DAKO) was applied and allowed to react at room temperature for 30 min. Following a wash with pH 7.6 Tris-HCl buffer, peroxidase-labeled streptavidin solution (100 μL) was applied and allowed to react at room temperature for 15 min. The sections were washed again with pH 7.6 Tris-HCl buffer, air-dried, counterstained with hematoxylin, and mounted with Permount (Thermo Fisher Scientific, Waltham, MA, USA) for microscopic examination.

Interpretation of IHC staining involved classifying the extent of staining on a scale of 0 to 3 (0, none; 1+, ≤ 10%; 2+, 10% to 25%; 3+, > 25%) for each protein antigen.

2.4. Next-generation sequencing

Genetic analysis using NGS was conducted to discover gene mutations and new markers that induce tumors within cancer tissues confirmed as lung cancer. Initially, both unstained 10 μm FFPE slides and 4 μm H&E-stained slides were prepared. Tumor regions were marked on the H&E slides using a fine-tip pen or an oil-based pen, corresponding to the tumor areas on the unstained slide. After marking, the digestion buffer was applied to the tissue sections, and the tissue was scraped off and transferred to tubes. DNA and ribonucleic acid (RNA) were then extracted using the RecoverAll Multi Sample RNA/DNA Kit (Thermo Fisher Scientific).

DNA concentration was measured using the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific), and RNA concentration was measured using the Qubit RNA HSAssay Kit (Thermo Fisher Scientific), both using the Qubit 4 Fluorometer (Thermo Fisher Scientific). RNA was reverse transcribed into complementary DNA (cDNA) using the SuperScript IV VILO Master Mix (Thermo Fisher Scientific). The cDNA (TaqMan Reagent StarterKit, Thermo Fisher Scientific, MA, USA) and DNA (TaqMan RNase P Detection Kit; Thermo Fisher Scientific) were then quantified by real-time PCR and sequenced using the Ion S5 XL (Thermo Fisher Scientific) instrument.

Data analysis was performed using bioinformatics tools, including the Torrent Mapping Alignment Program and the Torrent Variant calling algorithm provided by Ion Torrent (Thermo Fisher Scientific). The criteria for variant calling interpretation are outlined in **Table 1**.

Gene mutation evaluation in this study was based on the widely used four-tier evaluation system worldwide. Specifically, gene mutations with strong clinical significance (Tier I) or potential clinical significance (Tier II) were selected for application.

2.5. Real-time polymerase chain reaction for *EGFR* mutation detection

To ensure the accuracy of NGS results, *EGFR* real-time PCR was performed, and the results were utilized to assess accuracy and precision. First, DNA was extracted from 10 μm FFPE sections using the QIAamp DNA Mini Kit (Qiagen, Hilden, Germany). The concentration of the extracted DNA was measured using the NanoDrop 2000 (Thermo Fisher Scientific). Subsequently, the EGFR mutation genotype was detected using the PANA Mutyper™ R *EGFR* Kit (Panagene, Daejeon, Korea), and PCR was performed using the CFX96 Touch Real-Time PCR Detection System (BIO-RAD, Hercules, CA, USA).

In this process, wild-type complementary peptide nucleic acid (PNA) probes were added to suppress the amplification of wild-type sequences, selectively amplifying the small amount of *EGFR* mutation. The real-time PCR procedure included 15 cycles of denaturation step at 95℃ for 30 seconds, followed by 20 seconds of PNA clamping at 70℃, and annealing at 63℃ for 1 minute. Subsequently, amplification was carried out through 35 cycles, involving denaturation at 95℃ for 10 seconds, annealing and detection at 53℃ for 20 seconds, and extension at 73℃ for 20 seconds.

Table 1. Next-generation sequencing variant call criteria

Nucleotide variation	Variant call criteria
SNV/INDEL	VAF: ≥ 5% in SNV
	5% in INDEL
CNV	Amplification: average CNV ≥ 4 (gain), < 1 (loss)
Translocation (fusion gene)	Read counts: ≥ 20
	Total valid mapped reads: ≥ 50,000

Abbreviations: SNV/INDEL, single nucleotide variation/ small insertions, deletion; VAF, variation allele frequency; CNV, copy number variation

The results were compared with the NGS results obtained from the same clinical samples.

2.6. Fluorescence *in situ* hybridization for *ROS1* mutation detection

The FISH-*ROS1* testing was conducted to ensure the accurate analysis of NGS results and to evaluate the concordance.

Two FFPE sections, cut to a thickness of 1–2 μm, were prepared, mounted on slides, and air-dried. Using one of the tissue sections, H&E staining was performed, and the tumor area was marked on the H&E slide with a fine-tip pen or an oil-based pen. Corresponding areas on the unstained slide were also marked. Afterward, incubation was performed at 56℃–60℃ for 2 hours in the Hybrite (Bimedis, Osceola, Olando, USA). The slides were then deparaffinized twice for 5 minutes in xylene and processed three times for 5 minutes in 100% ethanol to remove xylene. Subsequently, the slides were air-dried for 5 minutes at room temperature. They were then reacted in an 80°C pretreatment solution for 30 minutes, rinsed twice in distilled water for 2 minutes, and blotted dry with a paper towel.

Following this, the ZytoLight SPEC *ROS1* Dual Color Break Apart Probe (Zytovision, Bremerhaven, Germany) was used for the FISH-*ROS1*. Slides were placed at 37℃, and pepsin solution was applied for 16 minutes. Then, a 1% formaldehyde solution was applied at room temperature for 5 minutes, followed by treatment with wash buffer saline-sodium citrate at room temperature for 5 minutes and a rinse in distilled water for 1 minute. The slides were subsequentially dehydrated in 70%, 85%, and 100% ethanol and air-dried at room temperature for 5 minutes. The area where the *ROS1* probe was to be placed was then marked on the back of the slide with an oil-based pen, and approximately 5 μL of the *ROS1* probe was placed on a cover glass in the dark, covered well, and sealed with a paper bond. The slides were then placed on a Thermobrite (Leica Biosystems, Barrington, IL, USA) and hybridized for 14 to 24 hours. The slides were then washed twice with wash buffer A for 5 minutes at 37°C, dehydrated in 70%, 85%, and 100% ethanol sequentially, and air-dried at room temperature while protected from light.

Nuclear staining was performed using 4’,6-diamidino-2-phenylindole (DAPI). The probe’s target for hybridization is on the opposite side of the *ROS1* gene’s breakpoint. The 3’ *ROS1* probe hybridizes toward the terminal end from the breakpoint and spans approximately 215 kb with SpectriumOrange fluorescence. The 5’ *ROS1* probe hybridizes toward the centromere side from the breakpoint and spans approximately 715 kb with SpectrumGreen fluorescence.

The positive criteria for FISH-*ROS1* were established as follows: at least 15% of the observed cancer cells show a separation of orange and green signals with a distance of more than one signal diameter among 50 nuclei or the presence of only fused signals and green signals in more than 15% of the observed cancer cells. Negative criteria were defined as 2 or more fused signals or 2 or more orange signals in fused signals in more than 15% of the observed cancer cells. If 5 or fewer cells among 50 cells are positive (5/50 or < 10%), it is considered negative. If 25 or more cells among 50 cells are positive (25/50 or > 50%), it is considered positive.

2.7. Immunohistochemistry for *ALK* mutation detection (D5F3)

To ensure the accurate analysis of NGS results and for concordance assessment, IHC-*ALK* (D5F3) testing was conducted. 4 μm FFPE sections were mounted on positively-charged glass slides and stained using the BenchMArk XT Automated Slide Stainer (Ventana Medical System, Tucson, AZ, USA). The staining was performed using the Optiview DAB IHC Detection Kit and Optiview Amplification Kit (Ventana Medical Systems). Immunostaining was carried out using the

Anti-IHC-*ALK* (D5F3) Rabbit Monoclonal Primary Antibody. To aid in result interpretation, a negative control immunoglobulin G (IgG) slide was created using Rabbit Monoclonal Negative Control IgG (Ventana Medical System) that matched all samples. Hematoxylin II was used for nuclear counterstaining, followed by post-counterstaining, and slides were mounted using Permount. Interpretation of IHC-*ALK* (D5F3) staining was done under a microscope according to the guidelines for the test. The interpretation of IHC staining for each protein antigen was categorized into 0–3 points (0, none; 1+, ≤ 10%; 2+, 10% to 25%; 3+, ≥ 25%) based on the range of staining observed in the nuclei and cytoplasm.

2.8. Statistical analysis

For statistical analysis in this study, SPSS software version 25.0 (IBM, Armonk, NY, USA) was utilized. Cross-tabulation (Pearson's chi-squared test) was performed to determine the associations between smoking and *EGFR* mutations, as well as smoking and *TP53* mutations. To evaluate the concordance of gene mutation diagnoses (*EFGR*, *ALK*, *ROS*) based on the testing method, Cohen's Kappa coefficient was calculated. A statistical significance level of *P*-value less than 0.05 was considered significant.

3. Result

3.1. Classification of lung adenocarcinoma tissues using H&E staining

FFPE samples, derived from resected lung adenocarcinoma tissues, were subjected to H&E staining, which allowed us to differentiate lung adenocarcinoma from other non-small cell carcinoma types based on morphological characteristics. The well-differentiated tumor tissues displayed lepidic, acinar, or papillary patterns (**Figure 1**). Among the observed patterns, the acinar pattern was the most common,

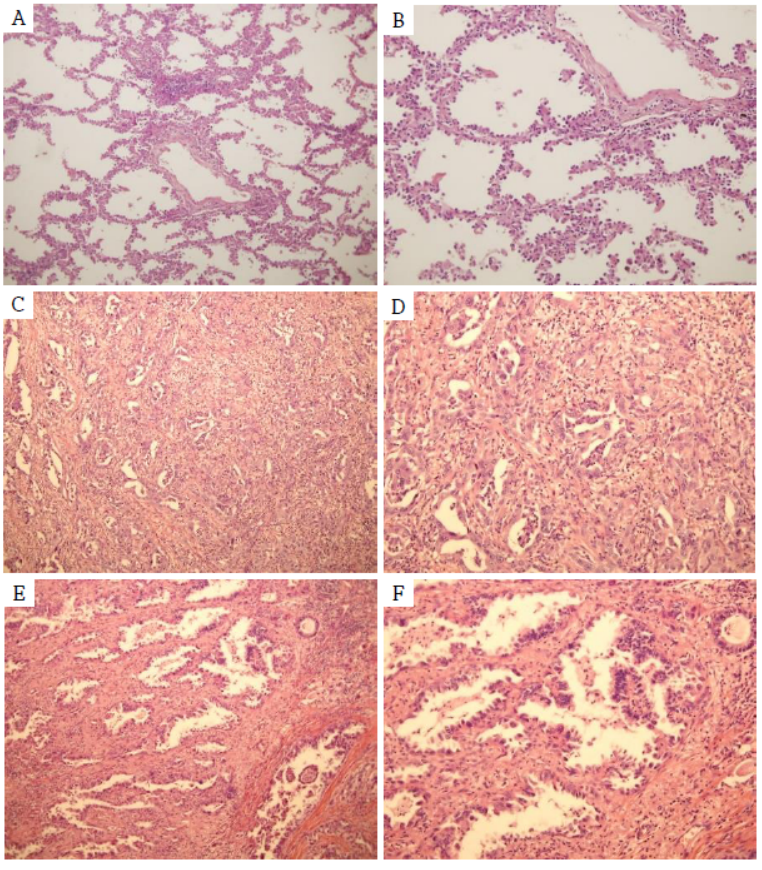

Figure 1. Subclassification of non-small cell lung cancer using paraffin section hematoxylin & eosin stain (A, C, E: 100×; B, D, F: 200×). **(A~F)**, well-differentiated lung adenocarcinoma (A, B: lepidic pattern; C, D: acinar pattern; E, F: papillary pattern).

accounting for 27 cases (32.5%), followed by the papillary pattern in 9 cases (10.8%), and the lepidic pattern in 7 cases (8.4%).

3.2. IHC staining to improve sensitivity and specificity of lung adenocarcinoma classification

In cases where tumor tissues were biopsy specimens, making it difficult to confirm the precise histological features of the tumor using H&E staining alone due to the scarcity of tumor cells or small tissue size, we employed IHC markers, including TTF-1, Napsin A, and P40, as primary antibodies for IHC staining to aid in lung adenocarcinoma classification [14]. Lung adenocarcinoma was discriminated by the presence of positive reactions in TTF-1 and Napsin-A staining and a negative reaction in P40 staining (**Figure 2**). As a result, we were able to confirm 26 cases (23%) of lung adenocarcinomas that were challenging to classify accurately based on H&E staining among non-small cell lung adenocarcinomas.

3.3. IHC staining for discriminating poorly differentiated lung adenocarcinoma

In cases of poorly differentiated lung adenocarcinoma, the combination of TTF-1, Napsin A, and P40 markers alone may not effectively distinguish lung adenocarcinoma from other non-small cell carcinomas. Therefore, we used CK7 and CK5/6 to further differentiate lung adenocarcinoma. Poorly differentiated lung adenocarcinoma was characterized by a negative reaction in P40 staining, weak positive reactions in TTF-1 and Napsin A staining, but a strong positive reaction in CK7 (**Figure 3**). As a result, we identified an additional 14 cases (12.5%) of poorly differentiated lung adenocarcinoma.

3.4. Clinical characteristics of non-small cell lung adenocarcinoma patients

To understand the clinical characteristics of all lung adenocarcinoma patients, a total of 83 patients were analyzed (**Table 2**). The cohort consisted of 34 males

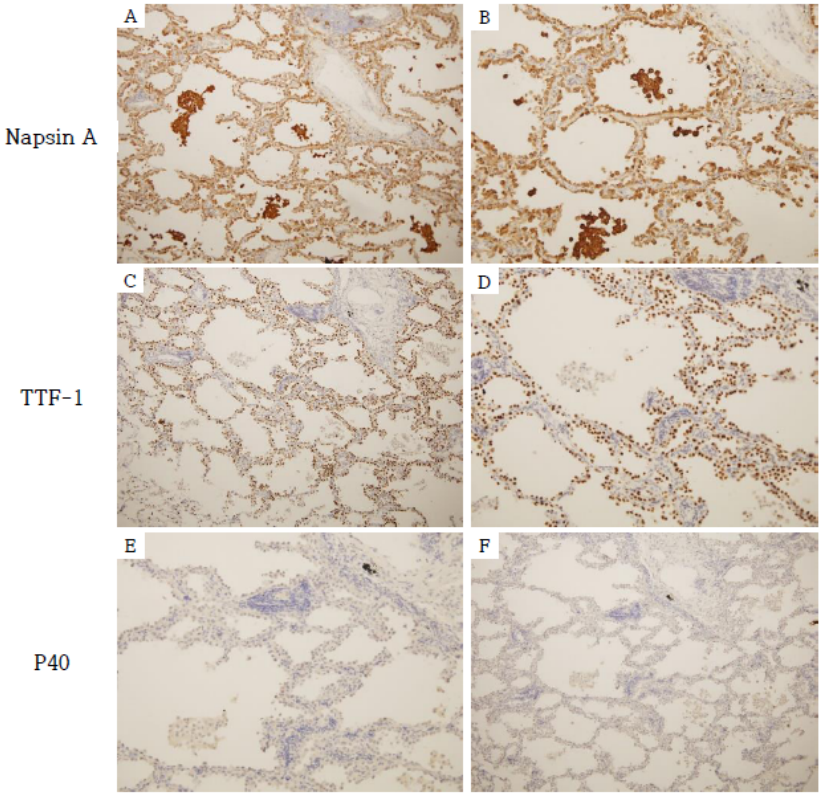

Figure 2. Subclassification of lung adenocarcinoma using immunohistochemical staining (A, C, E: 100×; B, D, F: 200×). Napsin A (+) and TTF-1 (+) are classified as lung adenocarcinoma. P40 (-) are not stained as lung adenocarcinoma carcinoma cells. **(A, B)** Napsin A (+); **(C, D)** TTF-1 (+); **(E, F)** P40 (-).

(41%) and 49 females (59%). Current or past smokers, defined as those with a smoking history of one year or more, constituted 34 patients (41%), while non-smokers comprised 49 patients (59%). The incidence rate of lung adenocarcinoma was higher among non-smokers (n = 49, 59%) compared to smokers (n = 34, 41%).

3.5. Analysis of SNVs in lung adenocarcinoma patients

Among the 83 patients confirmed as lung adenocarcinoma based on H&E and IHC staining, NGS analysis revealed that nearly 99% had at least one mutation. In total, 178 SNVs, including missense, nonsense, in-frame indels, and frameshift indels, were

Table 2. Baseline characteristics of analytic patients

Characteristics	Analytic patient (n = 83)
Lung adenocarcinoma	
Age, yr	65 (40–89)
Gender	
Male	34 (41)
Female	49 (59)
Smoking	
Former and current	34 (41)
Never	49 (59)

Data are presented as median (range) or number (%).

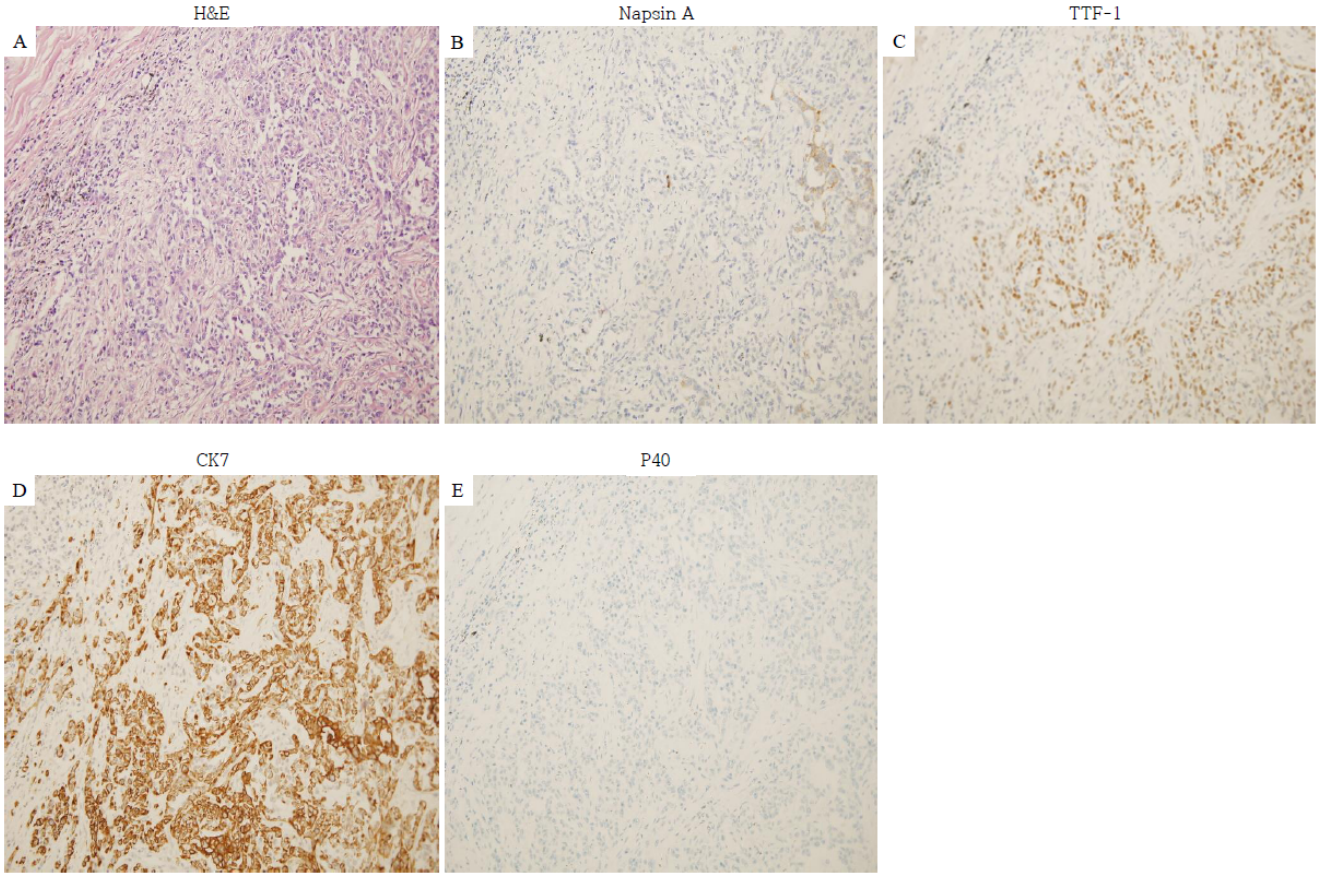

Figure 3. Results of H&E staining and immunohistochemistry staining of poorly differentiated lung adenocarcinoma (200×). **(A)** H&E; **(B)** Napsin A (weakly +); **(C)** TTF-1 (weakly +); **(D)** CK7 (strong +); **€** P40 (-) are classified as lung adenocarcinoma. Abbreviation: H&E, hematoxylin and eosin.

identified in 26 genes. The number of mutations per patient ranged from one to seven. The most frequently observed mutations in female patients were in *EGFR* (30, 61%) and *TP53* (30, 61%), followed by *PIK3CA* (5, 10%), *KRAS* (3, 6%), *POLE* (3, 6%), *CDKN2A* (2, 4%), *BRCA2* (2, 4%), *BRAF* (2, 4%), *ARID1A* (1, 2%), *FGFR4* (1, 2%), *IDH* (1, 2%), *KNSTRN* (1, 2%), *NOTCH1* (1, 2%), *MSH6* (1, 2%), *KIT* (1, 2%), *NF1* (1, 2%), *PALB2* (1, 2%), *SMARCA4* (1, 2%), *STED2* (1, 2%), *TET2* (1, 2%), and *U2AF1* (1, 2%). In maple patients, *TP53* (20, 85%) mutations were the most common, followed by *EGFR* (10, 29%), *KRAS* (9, 26%), *ARID1A* (3, 9%), *CDKN2A* (3, 9%), *BRCA2* (2, 6%), *ERBB2* (2, 6%), *KNSTRN* (2, 6%), *PIK3CA* (2, 6%), *POLE* (2, 6%), *TET2* (2, 6%), *ATRX* (1, 2%), *ATM* (1, 2%), *BRAF* (1, 2%), *CHEK1* (1, 2%), *MSH6* (1, 2%), *NOTCH1* (1, 2%), *STK11* (1, 2%), *U2AF1* (1, 2%), and *VHL* (1, 2%) (**Figure 4**).

EGFR and *TP53* mutations were the most common in Korean lung adenocarcinoma patients. The analysis of smoking history showed that *EGFR* mutations were significantly more frequent in non-smoking female patients (n = 29, 60.4%), whereas *TP53* mutations were prevalent in male smokers (n = 8, 26.7%). Overall, *EGFR* mutations were more frequent in non-smokers than in smokers (P = 0.007), while *TP53* mutations did not show significant differences between smoker and non-smoker groups (P = 0.539) (**Table 3**).

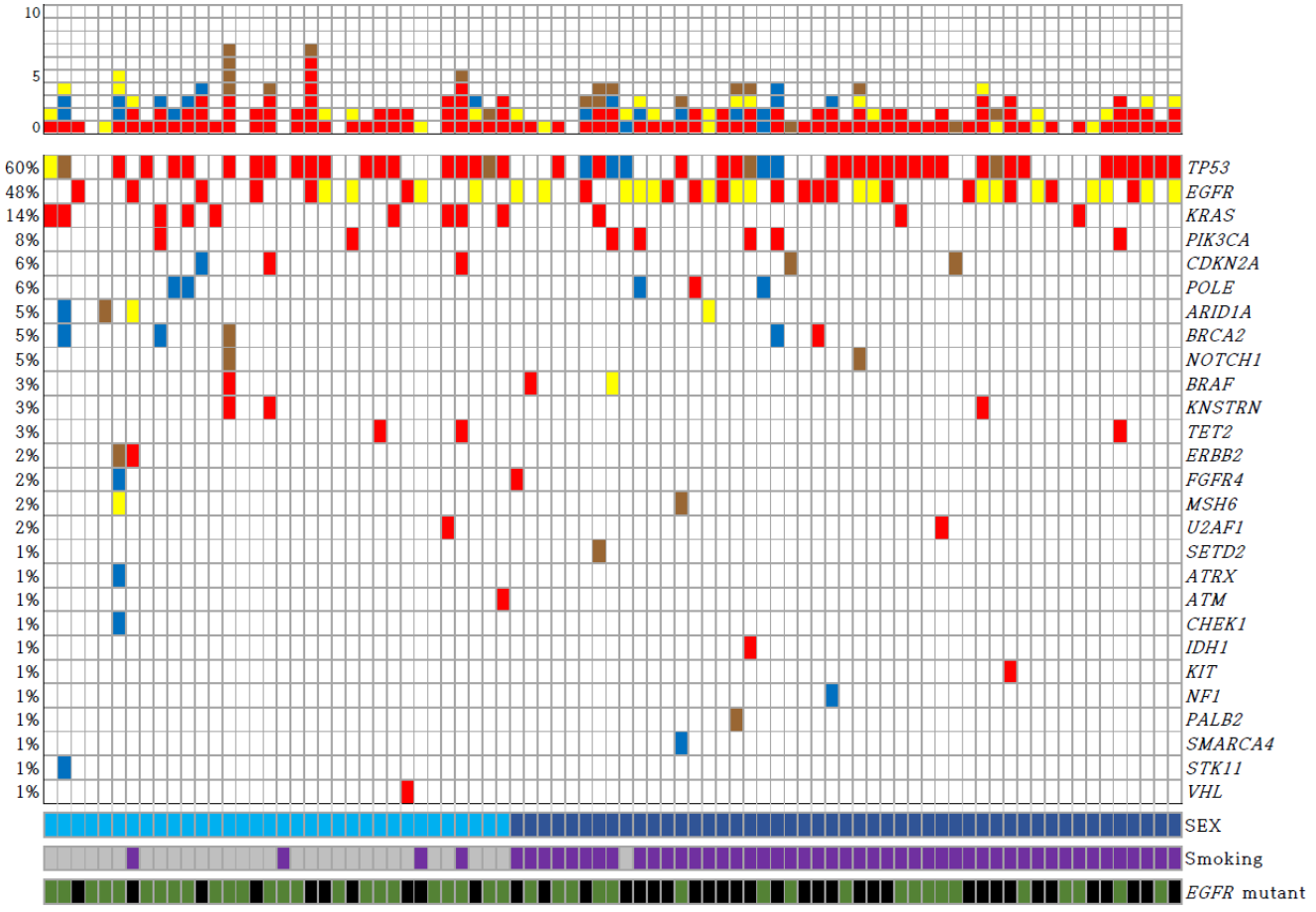

Figure 4. Next-generation sequencing analysis results, oncoprint heatmap of single nucleotide variations and insertions and deletions in lung adenocarcinoma. Missense is shown in red, nonsense in blue, in-frame indels in yellow, and frameshift indels in brown. The upper bar chart is the number of variations in each patient. The lower bar chart shows the gender and smoking status of each patient and the results of PCR for *EGFR*. Abbreviations: PCR, polymerase chain reaction; *EGFR*, epidermal growth factor receptor.

Table 3. Correlation between smoking and *EFGR* or *TP53* mutations

Variable	Smoking	Non-smoking
Female		
EGFR mutant	1 (100)	29 (60.4)
Wild-type	0 (0)	19 (36.6)
Male		
EGFR mutant	8 (26.7)	2 (50)
Wild-type	22 (73.3)	2 (50)
Total		
EGFR mutant	9 (29.0)	31 (59.6)
Wild-type	22 (71.0)	21 (40.4)
P-value	0.007	
Female		
TP53 mutant	1 (100)	29 (60.4)
Wild-type	0 (0)	19 (36.6)
Male		
TP53 mutant	19 (63.3)	1 (25.0)
Wild-type	11 (36.7)	3 (75.0)
Total		
TP53 mutant	20 (64.5)	30 (57.7)
Wild-type	11 (35.5)	22 (42.3)
P-value	0.539	

Data are presented as numbers (%). Abbreviation: *EGFR*, epidermal growth factor receptor.

3.6. CNV analysis in lung adenocarcinoma patients

A total of 132 CNVs were observed in 43 genes, and we confirmed the presence of both amplifications and deletions in CNVs. CNVs were detected in 41 of 83 lung adenocarcinoma patients (49%). Each patient had between 1 to 10 amplifications or deletions. Amplifications were more predominant, with 118 cases (89%), compared to 14 cases of deletions (11%) (**Figure 5**). *EGFR* amplifications were the most frequent among both female (6 cases, 12%) and male (6 cases, 14%) patients (total 12 cases, 10%). In patients with *EGFR* amplifications, 6 out of 6 female patients and 3 out of 6 male patients had co-occurring *EGFR* single nucleotide variations. Among these cases, *EGFR* exon 19 deletions were the most common, accounting for 7 cases (89%).

3.7. Gene rearrangement analysis in lung adenocarcinoma patients

Using NGS, we confirmed 10 gene fusions and 1 MET exon 14 skipping (**Table 4**). Among the gene fusions, *EML4-ALK* fusion (2, 2%) and *KIF5B-RET* fusion (2, 2%) were the most frequent, while other fusions, such as *TPM3-NTRK1*, *SDC4-ROS1*, *SLC34A2-ROS1*, *EZR-ROS1*, and MET exon 14 skipping, each occurring once. Among these, *ROS1* gene fusion was the most prevalent.

3.8. Concordance analysis between PCR and NGS results

To enhance the reliability of NGS results, we conducted PCR testing for *EGFR* hotspot mutations (exon 19 deletion, exon 20, exon 21 mutations) that are commonly found in East Asian female lung adenocarcinoma patients. We compared the results of PCR and NGS and confirmed a 100% (40/40) concordance (**Table 5**).

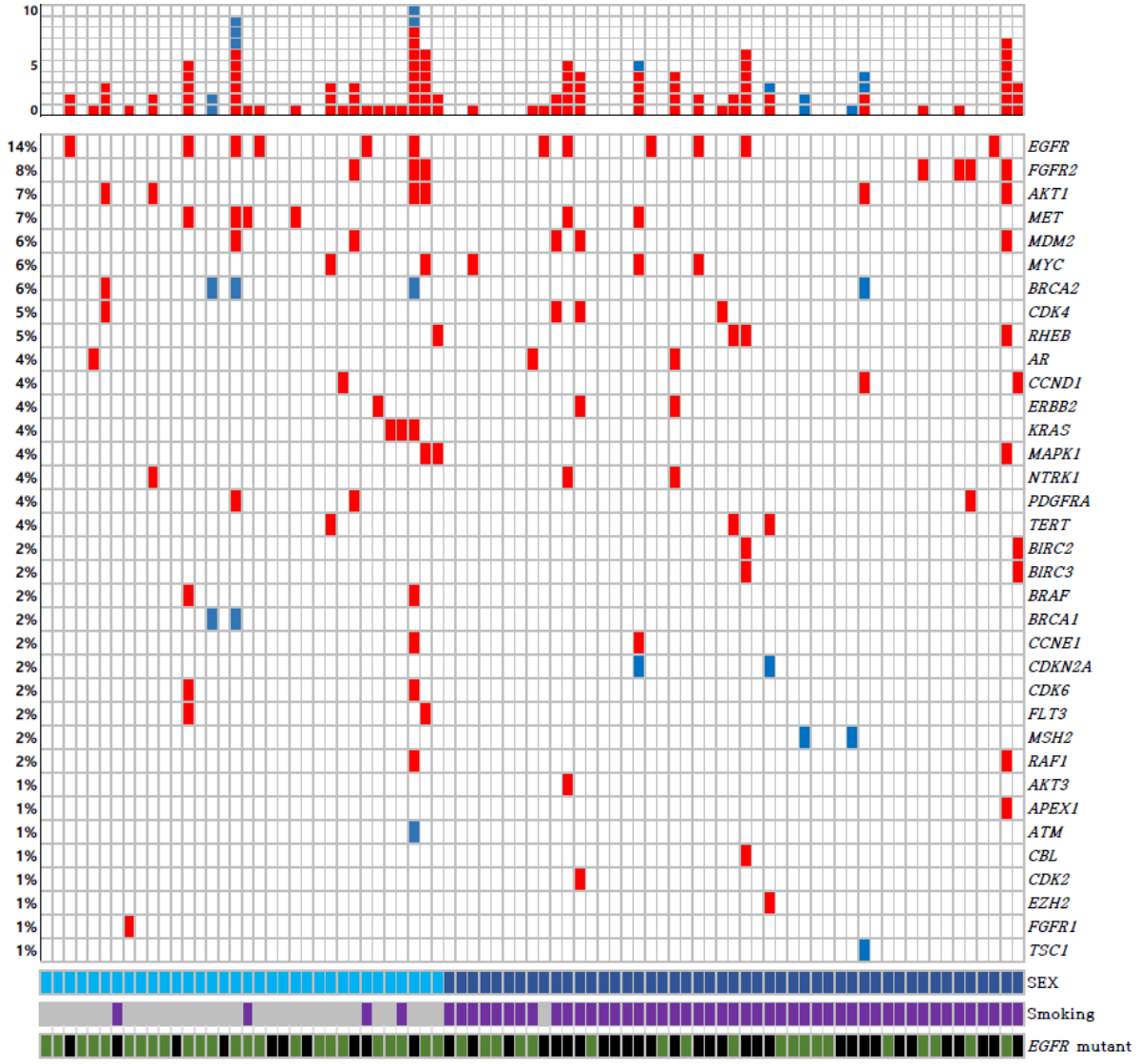

Figure 5. Next-generation sequencing analysis results, oncoprint heatmap of copy number variations in lung adenocarcinoma. Amplification is shown in red, and deletion in blue. The upper bar chart is the number of variations in each patient. The lower bar chart shows the gender and smoking status of each patient and the results of PCR for *EGFR*. Abbreviation: PCR, polymerase chain reaction; *EGFR*, epidermal growth factor receptor.

Table 4. Gene rearrangement of lung adenocarcinoma patients using next-generation sequencing analysis

Characteristics	Analytic patients (*n* = 83)
EML4-ALK fusion	2 (2)
KIF5B-RET fusion	2 (2)
TPM3-NTRK1 fusion	1 (1)
SDC4-ROS1 fusion	1 (1)
SLC34A2-ROS1 fusion	1 (1)
EZR-ROS1 fusion	1 (1)
RET-PCM1 fusion	1 (1)
MET exon 14 skipping	1 (1)

Data are presented as numbers (%).

Table 5. Comparison of PCR-*EGFR* test result and NGS result

Case No.	PCR-*EGFR*	NGS
1	*EGFR* exon 19 deletion	Glu746_Ala750del
3	*EGFR* exon 19 deletion	Leu747_Ala750delinsPro
6	T790M mutation in *EGFR* exon 20/L858R mutation in *EGFR* exon 21	Leu858Arg
9	*EGFR* exon 19 deletion	Glu746_Ala750del
10	*EGFR* exon 19 deletion	Leu747_Pro753delinsSer
11	*EGFR* exon 19 deletion	Glu746_Ala750del
12	L858R mutation in *EGFR* exon 21	Leu858Arg
14	L858R mutation in *EGFR* exon 21	Leu858Arg
15	*EGFR* exon 19 deletion	Glu746_Ala750del
16	L858R mutation in *EGFR* exon 21	Leu858Arg
17	*EGFR* exon 19 deletion	Glu746_Ala750del
18	*EGFR* exon 19 deletion	Leu747_Pro753delinsSer
20	L858R mutation in *EGFR* exon 21	Leu858Arg
22	L858R mutation in *EGFR* exon 21	Leu858Arg
23	L861Q mutation in *EGFR* exon 21	Leu861Gln
24	L858R mutation in *EGFR* exon 21	Leu858Arg
26	*EGFR* exon 19 deletion	Glu746_Ala750del
27	*EGFR* exon 19 deletion	Glu746_Ala750del
28	L858R mutation in *EGFR* exon 21	Leu858Arg
34	L858R mutation in *EGFR* exon 21	Leu858Arg
35	*EGFR* exon 19 deletion	Glu746_Ala750del
36	*EGFR* exon 19 deletion	Leu747Pro
37	L861Q mutation in *EGFR* exon 21	Leu861Gln
39	*EGFR* exon 19 deletion	Glu746_Arg748del/Thr751_Lys754delinslleSerProGlu
40	L858R mutation in *EGFR* exon 21	Leu858Arg
43	*EGFR* exon 19 deletion	LeuGlu746_Ser752delinsVal
44	*EGFR* exon 19 deletion	Glu746_Ala750del
46	L858R mutation in *EGFR* exon 21	Leu858Arg
47	*EGFR* exon 19 deletion	Leu747_Ser752del
49	*EGFR* exon 19 deletion	Glu746_Ala750del
52	L858R mutation in *EGFR* exon 21	Leu858Arg
56	L858R mutation in *EGFR* exon 21	Leu858Arg
61	*EGFR* exon 19 deletion/T790M mutation in *EGFR* exon 20	Glu746_Ala750del/Thr790Met/Cys797Gly
65	L858R mutation in *EGFR* exon 21	Leu858Arg
69	L858R mutation in *EGFR* exon 21	Leu858Arg
70		Glu709_Thr710delinsAsp
72	*EGFR* exon 19 deletion	Glu746_Ala750del
76	L858R mutation in *EGFR* exon 21	Leu858Arg
77	*EGFR* exon 19 deletion	Leu747_Thr751del
81	*EGFR* exon 19 deletion	Glu746_Ala750del

Abbreviations: PCR, polymerase chain reaction; *EGFR*, epidermal growth factor receptor; NGS, next-generation sequencing.

3.9. Concordance analysis between IHC-*ALK* and NGS results

IHC was used to perform *ALK* rearrangement tests on FFPE samples from 73 lung adenocarcinoma patients. The *ALK* rearrangement test was performed using IHC, which utilized the Ventana *ALK* (D5F3) companion diagnostic analysis method approved by the U.S. Food and Drug Administration to detect *ALK* protein expression (**Figure 6**) [15]. As a result, *ALK* rearrangements were observed in 3 cases of lung adenocarcinoma. To enhance the reliability of the results, we compared the outcomes of the Ventana *ALK* (D5F3) companion diagnostic method with NGS results. In NGS, two cases of *EML4-ALK* fusion were confirmed, and one case did not show any gene fusion.

3.10. Concordance analysis between FISH-*ROS1* and NGS results

We used FISH to observe *ROS1* gene rearrangement in FPPE samples from 77 lung adenocarcinoma patients. The FISH-*ROS1* test identified 3 cases of *ROS1* gene rearrangements (**Figure 7**). To analyze the concordance with NGS results, we compared the outcomes and found that 3 cases of *ROS1* gene fusions (*SDC4-ROS1* fusion, *SLC34A2-ROS1* fusion, and *EZR-ROS1* fusion) were also confirmed by NGS, indicating full concordance between FISH-*ROS1* and NGS results.

3.11. Concordance analysis between IHC-*ALK* (D5F3), FISH, PCR-*EGFR* testing, and NGS results

The results from previous IHC, FISH, and PCR tests were compared with NGS results. PCR-*EGFR* testing showed a 100% (κ = 1.00) concordance with the NGS results. The IHC-*ALK* (D5F3) testing had a concordance of 98.6% (κ = 0.79), while FISH-*ROS1* testing demonstrated 100% concordance (κ = 1.00) with NGS results (**Table 6**).

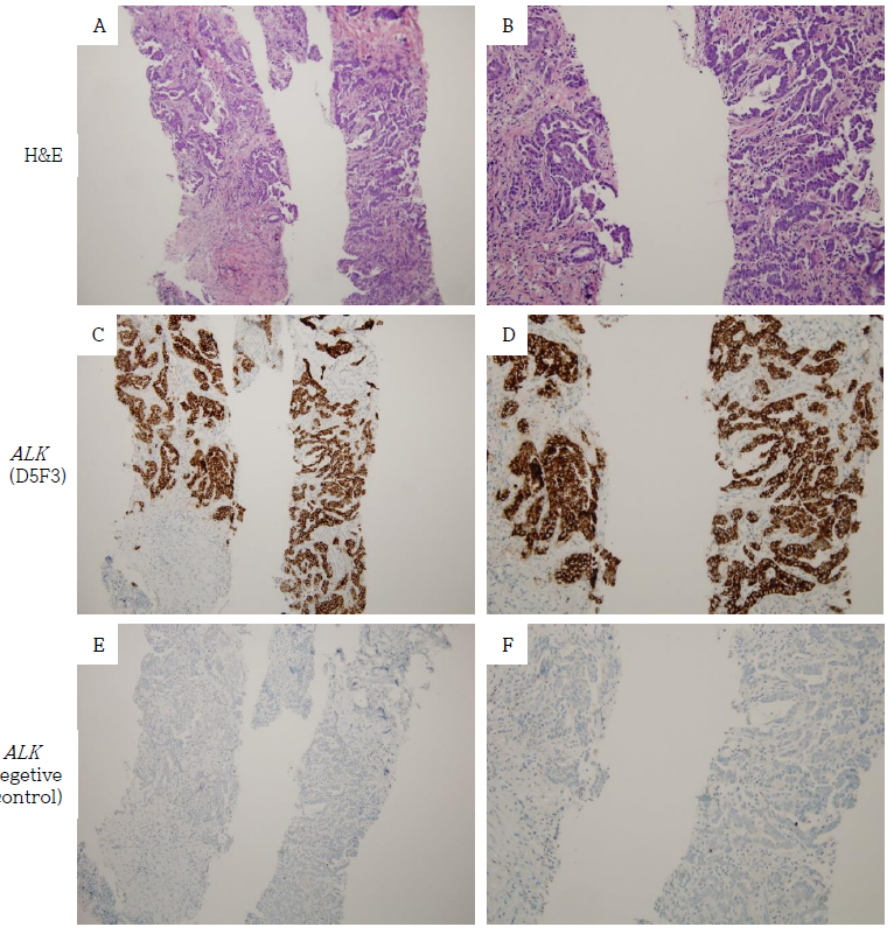

Figure 6. Results of IHC-*ALK* (D5F3) staining in lung adenocarcinoma tissue (A, C, E: 100×; B, D, F: 200×). **(A, B)** H&E stain; **(C, D)** positive IHC-*ALK* (D5F3); **(E, F)** negative IHC-*ALK* (D5F3) stain. Abbreviations: IHC, immunohistochemistry; *ALK*, anaplastic lymphoma kinase; H&E, hematoxylin and eosin.

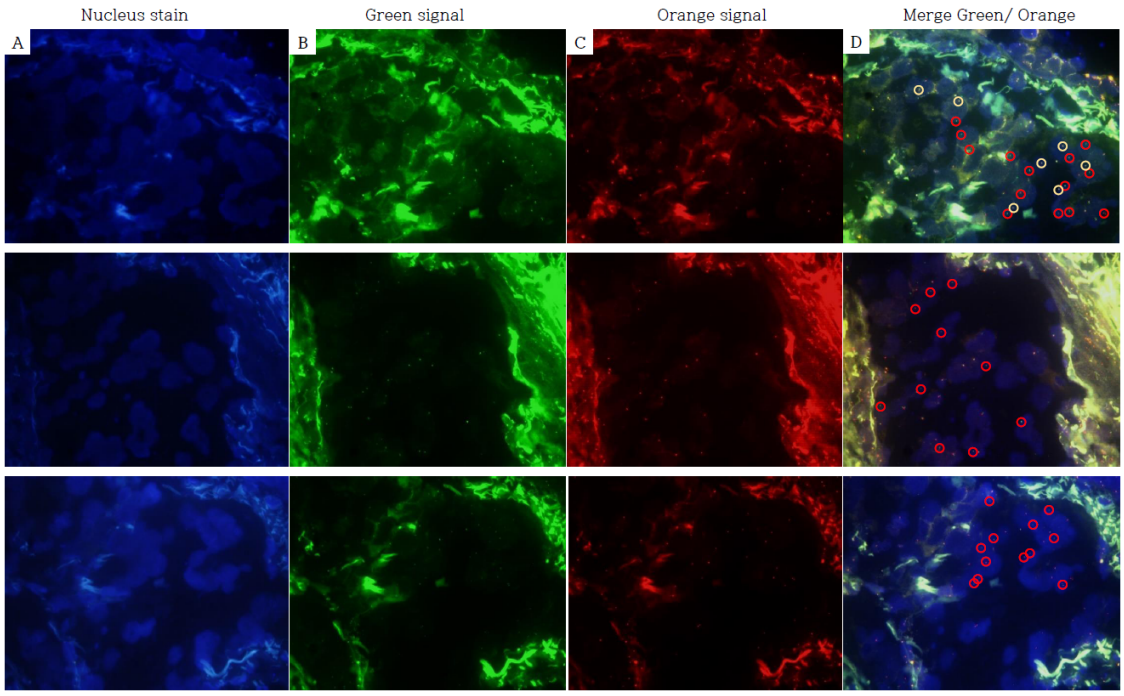

Figure 7. Result of FISH-*ROS1* test in lung adenocarcinoma tissue (1,000×). **(A)** Nuclei of tumor cells stained with DAPI; **(B)** Green signal detected through the green filter; **(C)** Orange signal detected through the green filter; **(D)** The result of merging the two signals. Red circles: *ROS1*-positive signal (split signal and single green signal); Yellow circles: *ROS1* negative signal (green/orange fusion signal and single orange signal). Abbreviation: *ROS1*: receptor tyrosine kinase 1; FISH, fluorescence *in situ* hybridization; DAPI, 4',6-diamidino-2-phenylindole.

Table 6. Concordance rates between PCR-*EGFR*, IHC-*ALK* (D5F3), and FISH-*ROS1* and NGS analysis in lung adenocarcinoma cases

Methodology	NGS		
	Positive	Negative	Total
PCR-*EGFR*			
Positive	40*	0	40
Negative	0†	43	43
Total	40	43	83
Concentration rates (%)		100 (κ = 1.00)	
IHC-*ALK* (D5F3)			
Positive	2	1	3
Negative	0	70	70
Total	2	71	73
Concentration rates (%)		98.6 (κ = 0.79)	
FISH-*ROS1*			
Positive	3	0	3
Negative	0	74	74
Total	3	77	77
Concentration rates (%)		100 (κ = 1.00)	

Abbreviations: PCR, polymerase chain reaction; *EGFR*, epidermal growth factor receptor; NGS, next-generation sequencing; IHC, immunohistochemistry; *ALK*, anaplastic lymphoma kinase; FISH, fluorescence *in situ* hybridization; *ROS1*, receptor tyrosine kinase 1. *In addition to the *EGFR* mutations, the two cases were accompanied by mutations at different locations in PCR (case 6), and NGS (case 11); †In one case, the *EGFR* mutation was confirmed in NGS, but it was located in exon 18 and judged to be accordant.

4. Discussion

Non-small cell lung cancer accounts for approximately 70% to 80% of all lung cancers and is further categorized into adenocarcinoma, squamous cell carcinoma, and large cell carcinoma. Among them, lung adenocarcinoma constitutes around 40% of all lung cancers [16]. Lung adenocarcinoma is associated with various genetic mutations, including *EGFR*, *KRAS*, *ALK*, *ROS1*, and others [11]. Particularly, lung adenocarcinoma driven by *EGFR* mutations, which are most prevalent in East Asian populations, has let the development of first-generation targeted therapies such as gefitinib and erlotinib, as well as second-generation agents such as afatinib and dacomitinib. Furthermore, third-generation targeted therapies have been developed for patients with resistance to first and second-generation treatments. Thus, the ability to detect *EGFR* mutations through NGS can offer significant clinical benefits. Targeted therapies have also been developed for *KRAS*, *ALK*, *MET*, and *ROS1* mutations, making the genetic analysis of lung adenocarcinoma crucial for patient treatment and prognosis, ultimately impacting survival rates [8].

The recent development of NGS has made it possible to analyze the genetic mutations in lung adenocarcinoma more rapidly and comprehensively by breaking down the entire genome into countless pieces and reading them simultaneously using bioinformatics techniques [9,10].

In this study, our focus was on the detection and frequency analysis of genetic mutations that can aid in the treatment of lung adenocarcinoma patients using NGS. We analyzed the relationship between the most common causative factor for lung cancer, smoking, and genetic mutations in lung adenocarcinoma, emphasizing the importance of NGS. To perform genetic analysis on lung adenocarcinoma tissue, it is essential to have accurate subclassification from the tissues of patients diagnosed with lung cancer. We followed the guidelines recommended by the World Health Organization, performing H&E staining, and further classifying lung adenocarcinoma tissues through IHC staining [14]. When lung adenocarcinoma patients were categorized by gender, age, and smoking status, a higher incidence was observed among females compared to males and among non-smokers compared to smokers, consistent with previous research findings [6].

The NGS results revealed that *TP53* (60.2%) mutations were the most common genetic variations in Korean lung adenocarcinoma patients. This result was consistent with the genetic mutation analysis of lung adenocarcinoma patients reported by The Cancer Genome Atlas Program [17]. However, *EGFR* mutations, which have been reported as the most frequent mutations in lung adenocarcinoma in East Asians [18-20], were not predominant in this study. In particular, *EGFR* was reported to be strongly associated with smoking [21], so we statistically analyzed the association of *EGFR* and *TP53* mutations with smoking in this study. *EGFR* mutations were primarily found in non-smokers (P = 0.007), while *TP53* mutations occurred regardless of smoking history (P = 0.539). These results suggest that the etiology of *TP53* mutations in lung adenocarcinoma among Koreans may be different and might be the major cause of lung adenocarcinoma in non-smoking females, but more comparative studies are needed to investigate the various factors that lead to *TP53* mutations and include a larger number of cases. Furthermore, NGS results indicated the presence of gene rearrangements (*ALK*, *ROS1*, *RET*, and *MET*) in lung adenocarcinoma patients, with the relative frequencies matching previous research findings [22].

To validate the reliability of the NGS results in this study, various additional tests were conducted. PCR-*EGFR*, IHC-*ALK* (D5F3), and FISH-*ROS1*, which are conventional genetic testing methods for diagnosing lung adenocarcinoma and are necessary for clinical patient care [23], were used to evaluate their concordance with NGS results. PCR-*EGFR* testing demonstrated 100% concordance with NGS results (κ = 1.00). The IHC-*ALK* (D5F3) testing had a case of discrepancy, resulting in 98.6% concordance (κ = 0.79), likely due

to limited tissue availability from small biopsy samples. FISH-*ROS1* testing showed 100% concordance with NGS results (κ = 1.00), highlighting that NGS is suitable for detecting various genetic mutations in lung adenocarcinoma in a single test.

In conclusion, using NGS for the diagnosis of genetic mutations in lung adenocarcinoma provides the advantage of simultaneously analyzing point mutations, insertions, deletions, fusions, rearrangements, CNV, and more. NGS has the potential to enable companion diagnostics for targeted therapies in lung adenocarcinoma and may help identify new biomarkers. This comprehensive approach is likely to be beneficial in clinical practice.

5. Conclusion

Lung cancer can be broadly categorized into small cell lung cancer and non-small cell lung cancer (NSCLC), with NSCLC representing approximately 70% to 80% of cases. Among NSCLC, lung adenocarcinoma accounts for about 40% of all lung cancers. Recent advancements in genetic profiling technologies have led to active research on mutations in important tumor genes and tumor suppressor genes related to tumor development and growth. Specific genes causing lung cancer have been discovered, significantly impacting survival rates. Particularly, in the case of lung adenocarcinoma, NGS has been instrumental in improving survival through companion diagnostics and targeted therapies. This study aimed to identify genetic mutations responsible for lung adenocarcinoma in Koreans. FFPE samples from NSCLC patients were used. Lung adenocarcinoma was distinguished using H&E staining, and further sub-classification was performed using IHC staining. Based on these results, NGS was employed to analyze the types and patterns of genetic mutations, and the relationship between lung cancer, especially lung adenocarcinoma, and the leading cause, smoking, was investigated. NGS results confirmed the presence of SNV, CNV, and gene rearrangements in lung adenocarcinoma. SNVs were observed in the following order: *TP53* (44. 6%), *EGFR* (35.7%), *KRAS* (10.7%), *PIK3CA* (6.2%), *CDKN2A* (4.4%). Among CNVs, *EGFR* (14%) was the most frequently occurring. Gene rearrangements, including *ALK*, *ROS1*, and *RET*, were also detected. To assess the reliability of NGS results, traditional genetic testing methods, such as PCR-*EGFR*, IHC-*ALK* (D5F3), and FISH-*ROS1*, were conducted additionally, and their concordance with NGS results was confirmed. This study highlights the positive impact of NGS in lung adenocarcinoma patients by simultaneously identifying mutations in multiple genes, enabling more favorable treatment strategies.

Disclosure statement

The authors declare no conflict of interest.

References

[1] Korea National Cancer Center, 2019, 2019 Cancer Registry Statistics Press Release, viewed, December 30, 2021, https://www.ncc.re.kr/cancerStatsView.ncc?bbsnum=578&searchKey= total&searchValue=&pageNum=1

[2] Union for International Cancer Control, 2018, Global Cancer Data: GLOBOCAN 2018, viewed, September 12, 2018, https://www.uicc.org/news/global-cancer-data-globocan-2018

[3] International Agency for Research on Cancer, 2022, Lung Cancer Awareness Month 2022, viewed, November 4, 2022, https://www.iarc.who.int/news-events/lung-cancer-awareness-month-2022/

[4] El-Telbany A, Ma PC, 2012, Cancer Genes in Lung Cancer: Racial Disparities: Are There Any? Genes & Cancer, 3(7–8): 467–480. https://doi.org/10.1177/1947601912465177

[5] Hirsch FR, Scagliotti GV, Mulshine JL, et al., 2017, Lung Cancer: Current Therapies and New Targeted Treatments. Lancet, 389(10066): 299–311. https://doi.org/10.1016/S0140-6736(16)30958-8

[6] Goldstraw P, Ball D, Jett JR, et al., 2011, Non-Small-Cell Lung Cancer. Lancet, 378(9804): 1727–1740. https://doi.org/10.1016/S0140-6736(10)62101-0

[7] Zappa C, Mousa SA, 2016, Non-Small Cell Lung Cancer: Current Treatment and Future Advances. Translational Lung Cancer Research, 5(3): 288–300. https://doi.org/10.21037/tlcr.2016.06.07

[8] Chan BA, Hughes BG, 2015, Targeted Therapy for Non-Small Cell Lung Cancer: Current Standards and the Promise of the Future. Translational Lung Cancer Research, 4(1): 36–54. https://doi.org/10.3978/j.issn.2218-6751.2014.05.01

[9] Van Allen EM, Wagle N, Stojanov P, et al., 2014, Whole-Exome Sequencing and Clinical Interpretation of Formalin-Fixed, Paraffin-Embedded Tumor Samples to Guide Precision Cancer Medicine. Nature Medicine, 20: 682–688. https://doi.org/10.1038/nm.3559

[10] U.S. Food and Drug Administration, 2017, Premarket Approval: Next Generation Sequencing Oncology Panel, Somatic or Germline Variant Detection System, viewed, June 24, 2019, https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpma/pma.cfm?id=p170019

[11] Desai A, Menon SP, Dy GK, 2016, Alterations in Genes Other than *EGFR/ALK/ROS1* in Non-Small Cell Lung Cancer: Trials and Treatment Options. Cancer Biology & Medicine, 13(1): 77–86. https://doi.org/10.28092/j.issn.2095-3941.2016.0008

[12] Tsimberidou AM, Elkin S, Dumanois R, et al., 2020, Clinical and Economic Value of Genetic Sequencing for Personalized Therapy in Non-Small-Cell Lung Cancer. Clinical Lung Cancer, 21(6): 477–481. https://doi.org/10.1016/j.cllc.2020.05.029

[13] Einaga N, Yoshida A, Noda H, et al., 2017, Assessment of the Quality of DNA from Various Formalin-Fixed Paraffin-Embedded (FFPE) Tissues and the Use of this DNA for Next-Generation Sequencing (NGS) with no Artifactual Mutation. PLoS ONE, 12(5): e0176280. https://doi.org/10.1371/journal.pone.0176280

[14] Osmani L, Askin F, Gabrielson E, et al., 2018, Current WHO Guidelines and the Critical Role of Immunohistochemical Markers in the Subclassification of Non-Small Cell Lung Carcinoma (NSCLC): Moving from Targeted Therapy to Immunotherapy. Seminars in Cancer Biology, 52(Pt 1): 103–109. https://doi.org/10.1016/j.semcancer.2017.11.019

[15] Uruga H, Mino-Kenudson M, 2018, *ALK* (D5F3) CDx: An Immunohistochemistry Assay to Identify *ALK*-Positive NSCLC Patients. Pharmacogenomics and Personalized Medicine, 2018: 147–155. https://doi.org/10.2147/PGPM.S156672

[16] Myers DJ, Wallen JM, 2013, Lung Adenocarcinoma, StatPearls Publishing, Treasure Island. https://www.ncbi.nlm.nih.gov/books/NBK519578/

[17] The Cancer Genome Atlas Research Network, 2014, Comprehensive Molecular Profiling of Lung Adenocarcinoma. Nature, 511: 543–550. https://doi.org/10.1038/nature13385

[18] Yatabe Y, Kerr KM, Utomo A, et al., 2015, *EGFR* Mutation Testing Practices within the Asia Pacific Region: Results of a Multicenter Diagnostic Survey. Journal of Thoracic Oncology, 10(3): 438–445. https://doi.org/10.1097/JTO.0000000000000422

[19] Kawaguchi T, Matsumura A, Fukai S, et al., 2010, Japanese Ethnicity Compared with Caucasian Ethnicity and Never-Smoking Status are Independent Favorable Prognostic Factors for Overall Survival in Non-Small Cell Lung Cancer: A Collaborative Epidemiologic Study of the National Hospital Organization Study Group for Lung Cancer

(NHSGLC) in Japan and a Southern California Regional Cancer Registry Databases. Journal of Thoracic Oncology, 5(7): 1001–1010. https://doi.org/10.1097/JTO.0b013e3181e2f607

[20] Chun YJ, Choi JW, Hong MH, et al., 2019, Molecular Characterization of Lung Adenocarcinoma from Korean Patients Using Next Generation Sequencing. PLoS ONE, 14(11): e0224379. https://doi.org/10.1371/journal.pone.0224379

[21] Husgafvel-Pursiainen K, Boffetta P, Kannio A, et al., 2000, p53 Mutations and Exposure to Environmental Tobacco Smoke in a Multicenter Study on Lung Cancer. Cancer Research, 60: 2906–2911.

[22] Kohno T, Nakaoku T, Tsuta K, et al., 2015, Beyond *ALK-RET*, *ROS1* and Other Oncogene Fusions in Lung Cancer. Translational Lung Cancer Research, 4(2): 156–164. https://doi.org/10.3978/j.issn.2218-6751.2014.11.11

[23] Cao B, Wei P, Liu Z, et al., 2015, Detection of Lung Adenocarcinoma with *ROS1* Rearrangement by IHC, FISH, and RT-PCR and Analysis of Its Clinicopathologic Features. OncoTargets and Therapy, 2016: 131–138. https://doi.org/10.2147/OTT.S94997