Mutation profile of non-small cell lung cancer revealed by next generation sequencing

Ya-Sian Chang  
China Medical University Hospital

Siang-Jyun Tu  
China Medical University Hospital

Yu-Chia Chen  
China Medical University Hospital

Ting-Yuan Liu  
China Medical University Hospital

Ya-Ting Lee  
China Medical University Hospital

Ju-Chen Yen  
China Medical University Hospital

Hsin-Yuan Fang  
China Medical University Hospital

Jan-Gowth Chang (✉ d6781@mail.cmuoh.org.tw)  
China Medical University Hospital  https://orcid.org/0000-0003-0375-1427

Research

Keywords: non-small cell lung cancer, whole-exome sequencing, targeted gene sequencing, trunk mutations

DOI: https://doi.org/10.21203/rs.3.rs-23075/v2

License: © This work is licensed under a Creative Commons Attribution 4.0 International License.  Read Full License
Abstract

Background: Precision therapy for lung cancer requires comprehensive genomic analyses. Specific effects of targeted therapies have been reported in Asian populations, including Taiwanese, but genomic studies have rarely been performed in these populations.

Method: We enrolled 72 patients with non-small cell lung cancer, of whom 61 had adenocarcinoma, 10 had squamous cell carcinoma, and 1 had combined adenocarcinoma and squamous cell carcinoma. Whole-exome or targeted gene sequencing was performed. To identify trunk mutations, we performed whole-exome sequencing in two tumor regions in four patients.

Results: Nineteen known driver mutations in EGFR, PIK3CA, KRAS, CTNNB1, and MET were identified in 34 of the 72 tumors evaluated (47.22%). A comparison with the Cancer Genome Atlas dataset showed that EGFR was mutated at a much higher frequency in our cohort than in Caucasians, whereas KRAS and TP53 mutations were found in only 5.56% and 25% of our Taiwanese patients, respectively. We also identified new mutations in ARID1A, ARID2, CDK12, CHEK2, GNAS, H3F3A, KDM6A, KMT2C, NOTCH1, RB1, RBM10, RUNX1, SETD2, SF3B1, SMAلة4، THRAP3，TP53，and ZMYM2. Moreover, all ClinVar pathogenic variants were trunk mutations present in two regions of a tumor. RNA sequencing revealed that the trunk or branch genes were expressed at similar levels among different tumor regions.

Conclusions: We identified novel variants potentially associated with lung cancer tumorigenesis. The specific mutation pattern in Taiwanese patients with non-small cell lung cancer may influence targeted therapies.

Background

Lung cancer is the leading cause of cancer-associated mortality worldwide. An estimated 2.09 million new lung cancer cases and 1.76 million lung cancer-associated mortalities were reported in the GLOBOCAN 2018 database [1]. Lung cancer can be divided into two broad categories according to histology: small-cell lung cancer and non-small-cell lung cancer (NSCLC). The latter comprises more than 80-85% of all lung cancers. NSCLC is subdivided into adenocarcinoma (ADC; 60%), squamous cell carcinoma (SqCC; 30-35%), large cell carcinoma, and other rare tumors, including adenosquamous carcinoma [2]. Risk factors include tobacco consumption, genetic susceptibility, poor diet, air pollution and occupational exposures, such as asbestos, metals and mixed occupation exposures, silica, polycyclic aromatic hydrocarbons, and diesel exhaust fumes [3]. Tobacco smoking is the main risk factor for lung cancer, however, 10-15% of those diagnosed with lung cancer are never-smokers [4]. Epidemiologic studies reveal that the proportion of lung cancer in never-smokers is higher in East Asia [5]. Never-smoker East Asian females are diagnosed more often with ADC, and these patients exhibit higher treatment response rates to epidermal growth factor receptor (EGFR) tyrosine kinase inhibitors [6]. Chest radiography, sputum cytology, and low-dose computed tomography (CT) have been used for lung cancer screening. Despite advances in genomic research and targeted therapies, leading to improvements in therapeutic strategies and the clinical outcomes of lung cancer patients [7], the overall 5-year survival rate of lung cancer remains very low (16.8%) [8]. The prognosis of lung cancer remains poor because most patients are often diagnosed at an advanced stage.

Rapid advancements in next-generation sequencing technology and a better understanding of cancer biology have provided unprecedented opportunities to characterize the genome of human tumors including lung cancer. The Cancer Genome Atlas (TCGA) lung cancer working group has profiled and analyzed 230 ADC and 178 SqCC specimens to identify molecular aberrations at the DNA, RNA, protein, and epigenetic levels [9, 10]. In ADC, the most common mutations are TP53, KRAS, EGFR, NF1, BRAF, MET, and RIT. Pathway alterations in ADC are involved in RTK/RAS/RAF, mTOR, JAK-STAT, DNA repair, and epigenetic deregulation. Mutations in TP53, CDKN2A, PIK3CA, NFE2L2, KEAP1, CUL3, PTEN, NF1, NOTCH1, 2 and 3, DDR2, and EGFR genes are frequently observed in SqCC. Pathway alterations in SqCC involve squamous differentiation, the oxidative stress response, PIK3CA, DNA repair, cell cycle regulation, and epigenetic deregulation [11]. ADC and SqCC show genetic alterations or gene expression differences [12, 13]. In ADC, the most common therapeutic targets are EGFR and BRAF mutations and ALK and ROS1 rearrangements. Molecular genotyping is now routine in ADC. In SqCC, targeted agents are largely ineffective, and many targeted therapies are currently undergoing clinical trials [14].

To develop a more comprehensive genomic picture of NSCLC, we performed whole-exome sequencing (WES) or targeted gene sequencing (TGS) in 72 Taiwanese patients with NSCLC. In addition, we compared the results with the TCGA NSCLC dataset, which involves mainly Western populations. We also investigated the associations between genetic alterations and clinicopathological features.

Materials And Methods

Subjects and DNA extraction

A total of 61 fresh-frozen and 11 formalin-fixed paraffin-embedded (FFPE) specimens were obtained from 72 Taiwanese patients with lung cancer who underwent surgical resection from May 2007 to April 2019 at the China Medical University Hospital. The 72 lung tumors comprised 61 ADCs, 10 SqCCs, and 1 combined ADC and SqCC. DNA from frozen tissues and FFPE specimens was extracted using the QIAamp® DNA Micro Kit (Qiagen, Heidelberg, Germany) and QIAamp® DNA FFPE Tissue Kit (Qiagen) according to the manufacturer’s instructions. The DNA concentration was quantified using the NanoDrop1000 spectrophotometer (Nanodrop Technologies, Wilmington, DE, USA) and a Qubit Fluorometer (Invitrogen, Carlsbad, CA, USA).

TGS and data analysis

TGS was performed using the Qiagen platform with a panel that included either 275 (cat. no. DHS-35012) or 72 genes (cat. no. DHS-0052). DNA libraries were prepared using components from the QiAseq Targeted DNA Panel Kit (Qiagen) and QiAseq Targeted DNA Panel Human Lung Cancer Panel (Qiagen). Briefly, 80 ng DNA was enzymatically fragmented and end-repaired in a 25 μl reaction volume containing 2.5 μl 10x fragmentation buffer and 5 μl fragmentation enzyme mix. The reaction was carried out at 4°C for 1 min, 32°C for 24 min, and 65°C for 30 min. Next, 10 μl 5x ligation buffer, 5 μl DNA ligase, and 2.8 μl 25
μM barcoded adapters were added along with enough water to reach a reaction volume of 50 μl. Reaction tubes were then incubated at 20°C for 15 min. To ensure complete removal of free barcoded adapters, each reaction was purified using 1.4x (or 1.0x) QIAseq beads for two rounds. The purified DNA was then mixed in a 20-μl reaction volume with 10 nM each target primer, 400 nM IL-Forward primer, 1x TEPCR buffer, and 0.8 μl HotStarTaq DNA polymerase. The PCR protocol was as follows: 95°C for 13 min; 98°C for 2 min; six cycles of 98°C for 15 s and 65°C for 15 min; and 72°C for 5 min. Each reaction was cleaned once using 1.4x (or 1.0x) QIAseq beads to remove unused primers. Enriched DNA was combined with 400 nM IL-index primers, 1x UPCR buffer, and 1 μl HotStarTaq DNA polymerase in a volume of 20 μl. The universal PCR conditions were as follows: 95°C for 13 min; 98°C for 2 min; 20 cycles of 98°C for 15 s and 60°C for 2 min; and 72°C for 5 min. The DNA library was purified once using 1.4x (or 1.0x) QIAseq beads and quantified using Qubit Fluorometric Quantitation (Thermo Fisher Scientific-US, Waltham, MA, USA). Libraries were sequenced on Illumina NextSeq (paired-end, 2 x 150 bp) according to the manufacturer’s instructions (Illumina, San Diego, CA, USA). TGS analysis was described in detail in our previous work [15].

**WES and data analysis**

A total of 50 ng DNA (based on Qubit quantitation) was tagmented by a transposome, followed by clean-up and amplification of the tagmented DNA. A 200-400 bp band was selected, and exome capture was performed using the Nextera Exome Library Preparation Kit (Illumina). The DNA library was quantified using the Qubit 3.0 Fluorometer (Invitrogen) and Agilent 4200 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). Samples were subjected to paired-end sequencing using the Illumina NovaSeq 6000 platform with a 150-bp read length. WES analysis has been described in detail in our previous work [16].

**RNA sequencing (RNA-seq)**

Total RNA was extracted from clinical tissue samples using the NucleoSpin® RNA Kit (MACHEREY-NAGEL GmgH, Düren, Germany) following the manufacturer’s instructions. The quality, quantity, and integrity of the total RNA were evaluated using the NanoDrop1000 spectrophotometer and Bioanalyzer 2100 (Agilent Technologies). Samples with an RNA integrity number >6.0 were used for RNA-seq. An mRNA-focused, barcoded library was generated using the TruSeq strand mRNA Library Preparation Kit (Illumina). The libraries were sequenced on the Illumina Nova Seq 6000 instrument (Illumina), using 2x151-bp paired-end sequencing flow cells following the manufacturer’s instructions.

**RNA-seq data analysis**

Illumina bcl2fastq Conversion Software (v2.20.0.422) was utilized to convert raw sequencing data to fastq format (Illumina). Trimmomatic PE (v0.39) was applied to control the read quality and remove sequencing adapters [17]. Reads were discarded if their average quality was <20 (AVGQUAL:20) and their read length <105 bp (MINLEN:105). Next, paired quality-controlled reads were aligned to the human genome (GRCh38), and gene expression was quantified using transcripts per million normalization via the HISAT2 (2.1.0) [18] and StringTie (1.3.5) [19] pipelines. To evaluate the similarities between different regions from the same tissue, we applied Spearman’s rank correlation of the transcripts per million values of 299 gene signatures [20]. A heatmap of the gene expression values was plotted using Morpheus (https://software.broadinstitute.org/morpheus), and the correlation coefficients were visualized using Seaborn, a Python data visualization library (https://github.com/mwaskom/seaborn).

**Statistical analysis**

All statistical analyses were performed using SPSS software ver. 22.0. Chi-squared or Fisher’s exact tests were used to compare two categorical variables. Survival analysis was performed using Kaplan-Meier survival plot and log-rank test. A p-value less than 0.05 was considered statistically significant.

**Results**

**Patient characteristics**

The patients in this study comprised 39 males and 33 females, with a mean age of 62.35 years, of whom 61 had ADC, 10 had SqCC, and 1 had combined ADC and SqCC. Stage I disease was identified in 38 patients, stage II in 12 patients, stage III in 10 patients, and stage IV in 6 patients (Table 1).

**Genomic alterations**

Among the 72 samples, 32 contained driver mutations in well-known cancer genes in NSCLC, such as *EGFR* (n=26; E709G, T790M, L858R and non-frameshift deletions of exon 19), and *PIK3CA* (n=4; E542K and G1049R). Besides *EGFR* and *PIK3CA*, other known mutations were detected in *KRAS* (n=4; G12V, G12A, G12D, and Q61H), which have all been reported as driver mutations in lung cancer. In addition, four samples carried known activating mutations in the well-known oncogenes *CTNNB1* (n=3; S33F, S37C and S37F) and *MET* (n=1; R1004X and c.3028+1G>T). Overall, 34 specimens harbored driver mutations in five cancer genes (*EGFR*, *PIK3CA*, *KRAS*, *CTNNB1*, and *MET*), which are canonical driver mutations (Table S1). These mutations were mutually exclusive, except for four cases of double mutations (n=2; *EGFR* and *CTNNB1* and n=2; *EGFR* and *PIK3CA*). *TP53* was the most frequently mutated gene after *EGFR* (n=18; 595fs, K120X, T125T, W146X, 152_153del, V173A, F212fs, G245C, G245D, R248L, R248W, R273H, R273C, V274D, E286K, and c.673-1G>T) (Table S1). Among the 16 *TP53* variants, two were novel (S95fs and F212fs). **Comparisons between Taiwanese and Caucasian patients with NSCLC**

To compare the frequency of driver mutations of NSCLC between Taiwanese and Caucasian patients, we obtained all available lung cancer cases (560 ADC and 489 SqCC) from the TCGA dataset. Notable differences from the TCGA data included the frequencies of mutations in *EGFR* (36.11% vs. 9.82%, p<0.0001), *KRAS* (5.56% vs. 15.92%, p=0.0165), and *TP53* (25.00% vs. 69.69%, p<0.0001). A full comparison of the frequencies of selected gene alterations between the two cohorts is depicted in Figure 1 and Table S2.

**Clinically relevant genomic alterations**
Based on the latest NSCLC guidelines published by the National Comprehensive Cancer Network, clinically relevant genomic alterations were identified in 34 (47.22%) patients (Table 2). As shown in Table 2, the clinically relevant alterations included in EGFR (26, 36.11%), ERBB2 (2, 2.78%), KRAS (4, 5.56%), MET (1, 1.38%), and NTRK1 (1, 1.38%).

Among the 26 patients with EGFR mutation patients, only 8 had an additional TP53 mutation, of whom 1 died of a cause unrelated to NSCLC. Among the 7 remaining patients, 3 had good survival outcomes, and 4 did not. We compared the genetic differences between the patients with a good and those with a poor survival outcome. In addition to the EGFR and TP53 mutations, one patient with poor survival harbored a MYC non-frameshift deletion (p.48_48del, rs776629119), and one patient with good survival had an AR non-frameshift insertion (p.L57delinsLQQQ, rs4045402) and a FBXW7 non-frameshift deletion (p.117_117del, rs781154022), and another patient with good survival had a CTNNB1 (p.S37F, rs121913403) mutation.

**Correlations between driver mutations and clinicopathological characteristics**

Correlations between the genotypes and clinicopathological characteristics are listed in Table 3. The EGFR mutation rate was significantly higher in patients with ADC than in those with SqCC (41.0% vs. 10.0%, \( p=0.059 \)). Moreover, the EGFR mutation rate was significantly higher in patients without smoking than in those with smoking (44.7% vs. 20.0%, \( p=0.038 \)). No association was found between the EGFR mutation status and sex, age, or tumor stage of the patients. In contrast, the PIK3CA and TP53 mutation rates were significantly higher in patients with SqCC than in those with ADC (30.0% vs. 1.6%, \( p=0.008 \) and 50.0% vs. 21.3%, \( p=0.053 \)). Furthermore, the TP53 mutation rate was significantly higher in patients with smoking than in those without smoking (40.0% vs. 17.0%, \( p=0.032 \)). No association was found between KRAS or CTNNB1 mutations and any clinicopathological characteristic.

We used Kaplan-Meier curve analysis to assess overall survival. In our cohort, PIK3CA mutation was a prognostic of worse overall survival (Figure 2). There was, however, no significant difference in mortality between EGFR, KRAS, TP53, and CTNNB1 mutations (Figure S1).

**Identification of trunk or branch driver mutations**

To examine intratumor heterogeneity, we applied multi-region WES in 8 tumor regions from 4 resected tissues (Figure 3 A). In order to determine whether driver genes carry trunk or branch mutations, we identified potential driver mutations among the 299 known cancer driver genes [20]. All variants classified as pathogenic in the ClinVar database are trunk mutations present in two tumor regions (Table S3). Among the 61 predicted pathogenic variants identified from four patients (15, 11, 16, and 19, respectively), 37 were classified as trunk mutations (11, 4, 7, and 15, respectively) (Figure 3 B, Table S4).

We further analyzed the variant allele frequencies of the trunk and branch mutations. Generally, the variant allele frequencies in four paired samples suggested that trunk mutations (median: 0.23-0.34%) occurred much more frequently than branch mutations (median: 0.12-0.15%) (Figure 3 C).

We also analyzed the expression of the driver genes that carried trunk or branch mutations. Gene expression profiles revealed no differences in driver genes harboring trunk or branch mutations between the two different tumor regions of the four paired samples (Figure 3 D).

**Intratumoral heterogeneity of 299 driver genes**

We determined the intratumoral expression of 299 driver genes, which were derived from 33 cancer types in the PanCancer dataset [20]. We used Spearman’s rank correlation to calculate the gene expression correlations between two regions from four tumors each. Two regions from a tumor showed the highest correlation coefficient (Figure 4).

The numbers of differentially expressed genes with a fold change in expression >4 in the four paired samples were 4 (FGFR2, PRKAR1A, MYC, and MYD88), 1 (MYD88), 4 (KMT2C, GNA11, ALB, and B2M), and 2 (KLFS and CDKN2A), respectively (Table S5). Most of the genes showed consistent expression (fold change ≤4). There were few differences between the different regions within the same tumor, and we suggest that any differences were due to branch mutations. Thus, the 299-driver gene signature may correctly predict cancer etiology if assessed from a single tumor region.

**Discussion**

In the present study, we successfully performed comprehensive genomic profiling in tumor specimens from 72 Taiwanese NSCLC patients using WES or TGS. We found that EGFR mutations were more common in patients with ADC, irrespective of sex, age, or tumor stage. PIK3CA and TP53 mutation rates were also higher in patients with SqCC. A comparison of driver gene mutations in our NSCLC patients with the TCGA dataset showed that EGFR was mutated at a much higher frequency in our Taiwanese cohort compared with Caucasians. In contrast, KRAS, BRAF, and TP53, the most common mutated genes in Caucasians, were found in only 5.56, 0, and 25%, respectively, of the Taiwanese NSCLC patients in our study, which is consistent with other studies [12, 21]. These differences are most likely due to racial and environmental differences.

In addition, we also identified 83 novel variants in 72 genes. Eighteen of these have been reported as cancer driver genes (ARID1A, ARID2, CDK12, CHEK2, GNAS, H3F3A, KDM6A, KMT2C, NOTCH1, RB1, RBM10, RYNX1, SETD2, SF3B1, SMARC4, THRAP3, TP53, and ZMYM2) (Table S6) [20]. Eight driver genes were associated with LUAD and LUSC.

**ARID1A** encodes a member of the SWI/SNF family of proteins. Mutation of ARID1A has been documented in a number of cancers [22], and approximately 8% of lung ADCs contain mutations in ARID1A [23]. In this study, we identified a novel ARID1A mutation (p.G933fs) in a patient with SqCC.

**KDM6A** is located on chromosome Xp11 and encodes a tetraticopeptide repeat protein. The protein contains a Jumonji C domain and catalyzes the demethylation of tri/dimethylated histone H3. KDM6A is a tumor suppressor gene in different cancers, including lung SqCC [20]. In this study, we identified a novel KDM6A mutation (p.S314X) in a patient with SqCC.
NOTCH1, a member of the NOTCH protein family, contains an extracellular domain consisting of multiple epidermal growth factor-like repeats (36) and an intracellular domain consisting of multiple domain types. NOTCH1 is reported to have a bimodal role as a tumor suppressor and an oncogene in several cancers [24, 25]. In this study, we identified a novel NOTCH1 mutation (p.E2534K) in a patient with SqCC.

RB1 encodes a negative regulator of the cell cycle protein and was the first tumor suppressor gene identified. Its tumor suppressive function is due to inhibition of the transcription factor E2F1. Any defect in the RB1 gene causes cells to transition from the G1 to S-phase of the cell cycle [26]. RB1 is inactivated in a wide range of cancers, including lung ADC and SqCC. In this study, we identified a novel RB1 mutation (p.I848fs) in a patient with ADC. The I848fs mutation occurs in the C domain, which mediates the interaction with E4F1 [27].

RBM10, a tumor suppressor gene [28, 29], encodes a nuclear protein that contains an RNA-recognition motif. RBM10 regulates pre-mRNA splicing in the alternative splicing pathway [30]. In this study, we identified a novel RBM10 mutation (p.A410fs) in a patient with ADC.

SETD2 encodes a protein that interacts with huntingtin. The protein is a histone methyltransferase responsible for the tri-methylation of lysine 36 on histone H3 (H3K36me3), using H3K36me2 as a substrate [31]. SETD2 is a tumor suppressor gene expressed in different cancer types [32]. In this study, we identified a novel SETD2 mutation (p.L1525P), present in the AWS domain, in a patient with SqCC.

SMARCA4, also known as BRG1, encodes a member of the SWI/SNF family, which possesses helicase and ATPase activities. Human SWI/SNF enzyme subunits are mutated in approximately 20% of cancers [33]. SMARCA4 is possibly tumor suppressive in lung ADC. In contrast, SMARCA4 may also be an oncogene in liver hepatocellular carcinoma, lower-grade glioma, and pan-cancer [20]. In this study, we identified a novel SMARCA4 mutation (p.Q570fs) in a patient with ADC.

Recently, Skoulidis et al. reported that co-occurring genomic alterations affect the response of NSCLC to anticancer therapies [34]. The mean overall survival was 82.30±9.80 months for the EGFR mutant TP53 wildtype cohort, 86.88±15.41 months for the EGFR/TP53 mutant cohort; p=0.839 (Figure S2). Among them, 11 patients used EGFR-TKI therapy. We also assessed whether there are differences in survival after EGFR-TKI therapy. The mean overall survival was 58.29±10.26 months for the EGFR mutant TP53 wildtype cohort (n=8), 114.50±2.50 months for the EGFR/TP53 mutant cohort (n=3); p=0.216 (Figure S3). The difference in survival time between these two groups was not significant, which may due to small sample size. We also assessed the overall survival of patients with EGFR/TP53-mutant lung cancers. Our cohort contained eight patients with co-existing EGFR/TP53 alterations. We compared the mutational signatures between the patients with a good and those with a poor survival outcome. The presence of a co-existing MYC alteration was associated with worse survival in patients with EGFR/TP53-mutant lung ADCs. Moreover, a co-existing AR, FBXW7, or CTNNB1 alteration was associated with better survival in patients with EGFR/TP53-mutant lung ADCs. In our study, a relatively small number of patients with co-existing EGFR/TP53 alterations was identified.

In summary, we identified genomic aberrations underlying NSCLC in a Taiwanese population. Our study provides putative biomarkers for prognostic prediction in lung cancer. Further research is required to elucidate the functions of these genes and their pathways.

Declarations

Acknowledgments

The data analyzed in this study were partially provided by the China Medical University Hospital Cancer Registry.

Authors’ contributions

YSC and JGC conceived and designed the experiments and wrote the paper. SJT performed computation analyses. YCC and TYL analyzed the data. YTL and JCY performed experiments. HYF collected the specimen and performed clinical review. JGC contributed reagents and materials and interpretation of data. All authors read and approved the final manuscript.

Funding

This work was supported by China Medical University Hospital (DMR-109-109 and DMR-109-248).

Ethics approval and consent to participate

The present study was approved by the Institutional Review Board of the China Medical University Hospital (CMUH106-REC1-053).

Consent for publication

Not applicable.

Competing interests
The authors declare that they have no competing interests.

References


Tables

**Table 1** Description of Taiwanese NSCLC cases.

<table>
<thead>
<tr>
<th>Variable</th>
<th>No of patients N=72 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years) Mean±SD</td>
<td>62.35±13.59</td>
</tr>
<tr>
<td>Range</td>
<td>36-83</td>
</tr>
<tr>
<td>Gender Male</td>
<td>39</td>
</tr>
<tr>
<td>Female</td>
<td>33</td>
</tr>
<tr>
<td>Clinical stage I</td>
<td>38 (55.78)</td>
</tr>
<tr>
<td>II</td>
<td>12 (16.67)</td>
</tr>
<tr>
<td>III</td>
<td>10 (13.89)</td>
</tr>
<tr>
<td>IV</td>
<td>6 (8.33)</td>
</tr>
<tr>
<td>Missing</td>
<td>6 (8.33)</td>
</tr>
<tr>
<td>Histology type</td>
<td></td>
</tr>
<tr>
<td>Adenocarcinoma</td>
<td>61 (84.72)</td>
</tr>
<tr>
<td>Squamous cell carcinoma</td>
<td>10 (13.89)</td>
</tr>
<tr>
<td>ADC+SqCC</td>
<td>1 (1.39)</td>
</tr>
<tr>
<td>Smoking status</td>
<td></td>
</tr>
<tr>
<td>Non-smokers</td>
<td>47 (65.28)</td>
</tr>
<tr>
<td>Smokers</td>
<td>25 (34.72)</td>
</tr>
</tbody>
</table>

**Table 2** Genomic alterations associated with targeted therapies.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Alteration</th>
<th>Targeted therapy</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any gene (s)</td>
<td></td>
<td>34</td>
<td></td>
</tr>
<tr>
<td><strong>EGFR</strong></td>
<td>L858R+E709G</td>
<td>Gefitinib</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Exon 19 deletion</td>
<td>Erlotinib, Gefitinib, Afatinib</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>L858R+T790M</td>
<td>Osimertinib</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>L858R</td>
<td>Erlotinib, Gefitinib, Afatinib</td>
<td>12</td>
</tr>
<tr>
<td><strong>KRAS</strong></td>
<td>G12V</td>
<td>Resistance to Erlotinib and Gefitinib</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>G12A</td>
<td>Resistance to Erlotinib and Gefitinib</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>G12D</td>
<td>Resistance to Erlotinib and Gefitinib</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Q61H</td>
<td>Resistance to Erlotinib and Gefitinib</td>
<td>1</td>
</tr>
<tr>
<td><strong>ERBB2</strong></td>
<td>Insertion</td>
<td>Afatinib, Dacomitinib, Trastuzumab</td>
<td>2</td>
</tr>
<tr>
<td><strong>MET</strong></td>
<td>R1004X</td>
<td>Crizotinib</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>c.1738+1G&gt;T</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>NTRK1</strong></td>
<td>R646C</td>
<td>Entrectinib</td>
<td>1</td>
</tr>
</tbody>
</table>

**Table 3** Correlation of **EGFR, PIK3CA, KRAS, TP53, and CTNNB1** mutations with clinicopathological features.
<table>
<thead>
<tr>
<th></th>
<th>EGFR mutation</th>
<th></th>
<th></th>
<th>PIK3CA mutation</th>
<th></th>
<th></th>
<th>KRAS mutation</th>
<th></th>
<th></th>
<th>TP53 mutation</th>
<th></th>
<th></th>
<th>CTNNB1 mutation</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No</td>
<td>Yes</td>
<td>Total</td>
<td>p-value</td>
<td>No</td>
<td>Yes</td>
<td>Total</td>
<td>p-value</td>
<td>No</td>
<td>Yes</td>
<td>Total</td>
<td>p-value</td>
<td>No</td>
<td>Yes</td>
<td>Total</td>
</tr>
<tr>
<td>Male</td>
<td></td>
<td></td>
<td>39</td>
<td>0.305</td>
<td></td>
<td></td>
<td>39</td>
<td>0.327</td>
<td></td>
<td></td>
<td>39</td>
<td>0.620</td>
<td></td>
<td></td>
<td>39</td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td></td>
<td>33</td>
<td>0.624</td>
<td></td>
<td></td>
<td>33</td>
<td>0.203</td>
<td></td>
<td></td>
<td>33</td>
<td>0.276</td>
<td></td>
<td></td>
<td>33</td>
</tr>
<tr>
<td>&lt;62</td>
<td></td>
<td></td>
<td>36</td>
<td>0.859</td>
<td></td>
<td></td>
<td>36</td>
<td>1.000</td>
<td></td>
<td></td>
<td>36</td>
<td>1.000</td>
<td></td>
<td></td>
<td>36</td>
</tr>
<tr>
<td>≥62</td>
<td></td>
<td></td>
<td>36</td>
<td>0.038</td>
<td></td>
<td></td>
<td>36</td>
<td>0.117</td>
<td></td>
<td></td>
<td>36</td>
<td>0.032</td>
<td></td>
<td></td>
<td>36</td>
</tr>
<tr>
<td>I-II</td>
<td></td>
<td></td>
<td>50</td>
<td>0.008</td>
<td></td>
<td></td>
<td>50</td>
<td>1.000</td>
<td></td>
<td></td>
<td>50</td>
<td>0.000</td>
<td></td>
<td></td>
<td>50</td>
</tr>
<tr>
<td>III-IV</td>
<td></td>
<td></td>
<td>16</td>
<td>1.000</td>
<td></td>
<td></td>
<td>16</td>
<td>1.000</td>
<td></td>
<td></td>
<td>16</td>
<td>1.000</td>
<td></td>
<td></td>
<td>16</td>
</tr>
<tr>
<td>ADC</td>
<td></td>
<td></td>
<td>61</td>
<td>0.000</td>
<td></td>
<td></td>
<td>61</td>
<td>1.000</td>
<td></td>
<td></td>
<td>61</td>
<td>0.000</td>
<td></td>
<td></td>
<td>61</td>
</tr>
<tr>
<td>SqCC</td>
<td></td>
<td></td>
<td>10</td>
<td>0.000</td>
<td></td>
<td></td>
<td>10</td>
<td>1.000</td>
<td></td>
<td></td>
<td>10</td>
<td>1.000</td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>Non-smokers</td>
<td></td>
<td></td>
<td>47</td>
<td>0.000</td>
<td></td>
<td></td>
<td>47</td>
<td>1.000</td>
<td></td>
<td></td>
<td>47</td>
<td>1.000</td>
<td></td>
<td></td>
<td>47</td>
</tr>
<tr>
<td>Smokers</td>
<td></td>
<td></td>
<td>25</td>
<td>0.000</td>
<td></td>
<td></td>
<td>25</td>
<td>1.000</td>
<td></td>
<td></td>
<td>25</td>
<td>1.000</td>
<td></td>
<td></td>
<td>25</td>
</tr>
</tbody>
</table>

p value by Chi-square or Fisher’s exact test when appropriated.