



OPEN

Comprehensive genomic profile of Chinese lung cancer patients and mutation characteristics of individuals resistant to icotinib/gefitinib

Yanhong Shang^{1,4}✉, Xiaofang Li^{1,4}, Weiwei Liu^{1,4}, Xiaoliang Shi², Shaohua Yuan², Ran Huo¹, Guotao Fang¹, Xiao Han¹, Jingnan Zhang¹, Kunjie Wang¹, Zhengyue Dou¹, Yan Zhang³, Aimin Zang¹ & Lin Zhang²

Lung cancer is the leading causes of cancer-related death worldwide. Precise treatment based on next-generation sequencing technology has shown advantages in the diagnosis and treatment of lung cancer. This cohort study included 371 lung cancer patients. The lung cancer subtype was related to the smoking status and sex of the patients. The most common mutated genes were *TP53* (62%), *EGFR* (55%), and *KRAS* (11%). The mutation frequencies of *EGFR*, *TP53*, *PIK3CA*, *NFE2L2*, *KMT2D*, *FGFR1*, *CCND1*, and *CDKN2A* were significantly different between lung adenocarcinoma and lung squamous cell carcinoma. We identified the age-associated mutations in *ALK*, *ERBB2*, *KMT2D*, *RBM10*, *NRAS*, *NF1*, *PIK3CA*, *MET*, *PBRM1*, *LRP2*, and *CDKN2B*; smoking-associated mutations in *CDKN2A*, *FAT1*, *FGFR1*, *NFE2L2*, *CCNE1*, *CCND1*, *SMARCA4*, *KEAP1*, *KMT2C*, and *STK11*; tumor stage-associated mutations in *ARFRP1*, *AURKA*, and *CBFB*; and sex-associated mutations in *EGFR*. Tumor mutational burden (TMB) is associated with tumor subtype, age, sex, and smoking status. TMB-associated mutations included *CDKN2A*, *LRP1B*, *LRP2*, *TP53*, and *EGFR*. *EGFR* amplification was commonly detected in patients with acquired icotinib/gefitinib resistance. *DNMT3A* and *NOTCH4* mutations may be associated with the benefit of icotinib/gefitinib treatment.

Lung cancer, the most common type of cancer, has the largest number of new cases and deaths worldwide¹. Non-small cell lung cancer (NSCLC), including lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC), is the most common subtype of lung cancer². The conventional treatment for lung cancer involves surgical resection, radiotherapy, and chemotherapy. Unfortunately, most patients are diagnosed at an advanced stage of the disease and miss the best treatment time, which leads to poor prognosis³. Early diagnosis and precise treatment are still the main obstacles in lung cancer treatment.

The continuous development of sequencing technology enables large-scale tumor-related gene detection. The mutational landscape of lung cancers, including LUSC and LUAD, has been reported by The Cancer Genome Atlas (TCGA)^{4,5}. Molecular targeted therapy and immunotherapy have the advantages of high accuracy, high conformability and fewer side effects^{6,7}. Ding et al. also compared the mutational features of LUSC and LUAD in Chinese NSCLC patients⁸. The understanding of molecular mutation characteristics in a large proportion of lung cancer patients may allow the development of personalized molecular targeted therapy or immunotherapy for patients with target mutations⁹.

Several genomic alterations that could be relevant in the clinical management of patients, such as *RET*, *ALK*, and *NTRK1* fusions, and *EGFR* and *KRAS* mutations, were identified and used for the exploration of targeted drugs^{10–17}. At present, targeted treatment of *EGFR* mutant tumors with *EGFR* tyrosine kinase inhibitors (TKIs)

¹Department of Medical Oncology, Affiliated Hospital of Hebei University, No. 648, Dongfeng East Road, Baoding City 071000, Hebei, People's Republic of China. ²Origimed Co. Ltd, Shanghai 201114, People's Republic of China. ³Department of Oncology, The First Hospital of Shijiazhuang City, No.36, Fanxi Road, Chang'an District, Shijiazhuang City 050000, Hebei, People's Republic of China. ⁴These authors contributed equally: Yanhong Shang, Xiaofang Li and Weiwei Liu. ✉email: hdfyzlnk@sina.com

has been used as the standard clinical treatment^{18–20}. However, cancer cells often develop resistance to these drugs, which may lead to progression²¹. Acquired EGFR T790M mutation is one reason for resistance to first- and second-generation EGFR-TKIs²². Osimertinib, a targeted drug to treat NSCLC patients with certain mutations demonstrated improved efficacy for EGFR T790M mutation²³. However, existing targets and targeted drugs are far from the cure for all lung cancer.

The different mutational characteristics of lung cancer patients from different regions affects the efficacy of target drugs. Satio et al. reported differences in driver gene aberration frequencies between Japanese and American patients⁶. This indicates different guiding effects of targeted drugs on the American and Asian populations. At present, most studies are focused on populations in Europe, America, Japan and South Korea^{10,13,15,16,24}. The study of the mutational profile of lung cancer patients in China may identify the genetic heterogeneity that is of great significance to the study of targeted therapy for Chinese patients.

In this study, we acquired 371 lung cancer tissue samples from Chinese patients at the Affiliated Hospital of Hebei University and determined their comprehensive genomic profiles to provide evidence for the potential development of special therapeutic targets and to identify new prognostic biomarkers.

Materials and methods

Ethics statement. This project was approved by the Ethics Committees of the Affiliated Hospital of Hebei University (approval number: HDFY-LL-2019-064). We declare that all methods used in this protocol were carried out in accordance with relevant guidelines and regulations. All patients and all participants provided informed consent.

Patients and sample collection. A total of 371 Chinese lung cancer patients were enrolled in this study from the Affiliated Hospital of Hebei University. Formalin-fixed paraffin-embedded (FFPE) tumor tissues and matched blood samples were collected and transferred to the laboratory of OrigiMed (Shanghai, China), which certified by College of American Pathologists and Clinical Laboratory Improvement Amendments, for genetic variation detection. Referring to previous methods^{25,26}, the genomic DNA was prepared using the QIAamp DNA FFPE Tissue Kit and QIAamp DNA Blood Midi Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. The concentration of DNA was measured by Qubit and normalized to 20–50 ng/μL for sequencing.

Identification of genomic alterations and tumor mutational burden. Genomic mutations were identified using the next-generation sequencing (NGS)-based YuanSu450 gene panel (OrigiMed, Shanghai, China), which covers all coding exons of 450 cancer-related genes that are frequently rearranged in solid tumors. The genes were captured and sequenced with a mean depth of 800× using Illumina NextSeq 500 (Illumina, Inc., San Diego, CA, USA). The procedures followed the steps described by Frampton et al.²⁷ Genomic alterations (GAs) were identified according to previous study²⁸. Single nucleotide variants (SNVs) were identified using MuTect (version 1.7, Broad Institute, Cambridge, MA, USA). Insertion-deletions (Indels) were identified using PINDEL (version 0.2.5). The functional impact of GAs was annotated using SnpEff3.0. Copy number variation regions were identified using Control-FREEC (version 9.7, Institute Cochin, Paris, France) with the following parameters: window = 50,000 and step = 10,000. Gene fusions were detected using an in-house developed pipeline. Gene rearrangements were assessed using Integrative Genomics Viewer. Tumor mutational burden (TMB) was estimated by counting the coding somatic mutations, including SNVs and Indels, per megabase of the sequence examined in each patient. The TMB value was further divided into two groups: TMB-H, defined as ≥ 10 mutations/Mb; TMB-L, defined as < 10 mutations/Mb.

Statistical analysis. Statistical analyses were performed using SPSS version 22.0 (SPSS Inc., Chicago, IL, USA). Fisher's exact test was used for the association analysis of categorical variables. The associations of normally and non-normally distributed data were analyzed using Student's *t*-test and Wilcoxon rank test respectively. The Kruskal–Wallis test was used to analyze the association between multiple groups of non-parametric data. *P* < 0.05 was considered statistically significant.

Results

Clinical characteristics of LC patients. The 371 Chinese lung cancer patients included 187 males and 184 females with a median age of 62 years (range, 27–84 years). The smoking status of 180 patients were collected, including 70 (18.87%) patients with a history of smoking > 10 years, and 110 (29.65%) who had never smoked. Among the smokers, 67 (95.7%) were male and 3 (4.3%) were female. Among non-smokers, 27 (24.5%) were male and 83 (75.5%) were female.

According to pathological classification, there were 335 (90.30%) NSCLCs, 11 (2.97%) small cell lung cancers (SCLCs), and 25 (6.74%) unclassified lung cancers in this cohort. The NSCLC consisted of 286 LUADs, 30 LUSCs, 8 LASCs, and 11 unclassified NSCLC (Table 1). Patients tumors were divided into stages I–IV based on the 8th edition Classification for lung cancer²⁹ and consisted of 56 stage I, 24 stage II, 50 stage III, 180 stage IV, and 61 unclear tumor stage samples (Table 1). Of all 371 samples used for NGS testing, 297 were from primary lesions, 56 were from metastases lesions—26 pleural (7.0%), 9 lymphatic (2.43%), 9 bone (2.43%), 3 liver (0.81%), 2 brain (0.54%), 2 peritoneal (0.54%), 1 adrenal gland (0.27%), and 4 unclear (1.08%)—and 18 samples were of unknown origin (Table 1).

Correlations between tumor subtype and smoking status, sex, tumor stage, and age of patients. Based on clinical information, we identified the correlation between tumor type and sex, smoking

Total		371
Gender (n/%)	Male	187 (50.40%)
	Female	184 (49.60%)
Age	Median (range)	62 (27–84)
Smoking status (n/%)	Smoking	70 (18.87%)
	Nonsmoking	110 (29.65%)
	Unknown	191 (51.48%)
TMB	Median (range)	3 (0–55.7)
Tumor sites (n/%)	Primary lesion	297 (80.05%)
	Metastases lesion	56 (15.09%)
	Pleural	26 (7.0%)
	Lymphatic	9 (2.43%)
	Bone	9 (2.43%)
	Liver	3 (0.81%)
	brain	2 (0.54%)
	peritoneal	2 (0.54%)
	Adrenal gland	1 (0.27%)
	Unclear	4 (1.08%)
	Unknown	18 (4.85%)
	Tumor stage (n/%)	I
II		30 (8.09%)
III		46 (12.40%)
IV		121 (32.61%)
Unknown		132 (35.58%)
Tumor subtype (n/%)	Non-small cell lung cancer	335 (90.30%)
	Lung adenocarcinoma	286
	Lung squamous cell carcinoma	30
	Lung adenosquamous carcinoma	8
	Unknown	11
	Small cell lung cancer	11 (2.97%)
	Unknown	25 (6.74%)

Table 1. Clinicopathologic features of this cohort. *TMB* tumor mutational burden.

history, age, and tumor stage. Considering the small number of SCLC and LASC cases, we excluded them from the correlation analysis. The proportion of patients with a smoking history was higher than those who never smoked in LUSC, while the proportion of patients who never smoked was higher than those with a smoking history in LUAD. Statistical analysis showed that smoking history was correlated with tumor subtype (Fig. 1A). In addition, we found that the proportion of male patients was higher than that of female patients with LUSC, while the proportion of female patients was higher than that of male patients with LUAD. Statistical analysis showed that the sex of patients was correlated with tumor subtype (Fig. 1B). Meanwhile, we found that the proportion of stage IV tumors was high in LUAD, while the proportion of stage I tumors was high in LUSC. Statistical analysis showed a significant association between tumor stage and tumor subtype (Fig. 1C). In addition, we found that the majority of patients were near 60 years old cross different tumor subtypes. Our results showed that there was no correlation between tumor subtype and patient age (Fig. 1D).

Genomic alterations in this cohort. A total of 2225 clinically relevant GAs were identified in this cohort, with a mean of 6.0 alterations per sample in 387 genes. Among these mutations, 1301 (68.5%) were substitution/indels, 487 (21.9%) were gene amplifications, 313 (14.1%) were truncations, 96 (4.3%) were fusions/rearrangements, and 28 (1.3%) were gene homozygous deletions (Fig. 2, Table S1). The most common mutated genes were *EGFR* (55%, 203/371), *TP53* (62%, 228/371), and *KRAS* (11%, 41/371). The most common gene fusion/rearrangement was *ALK*, and the main amplifications included *EGFR*, *PIK3CA*, *TERT*, *MET*, *SDHA*, *ERBB2*, *BRAF*, *NKX2-1*, *MDM2* and *CDK4* (Fig. 2, Table S1). In the same patient, one mutation site was detected in most genes, while two or three mutations were detected in a few genes such as *EGFR*, *TP53*, *KRAS*, *PIK3CA*, *ALK*, and *MET*.

In 203 lung cancer tissue samples, we detected a total of 313 *EGFR* GAs: 253 Substitution/Indels (including 250 Substitutions/Shortindels and 3 LongIndels, with 5 SNVs were belonging to germline mutation), 58 gene amplifications, and 2 fusions. The most common alteration was L858R, and followed by exon 19 deletion (19del) (Fig. S1).

The co-occurrence and mutual exclusion of gene mutations can influence prognosis. For this reason, we performed a co-mutation analysis of this cohort. Our results showed that mutations in *KMT2C*, *APC*, *CDKN2A*, *RBI*, and *EGFR* co-occurred with *TP53* mutations, while mutations in *MDM2* and *KRAS* were mutually exclusive

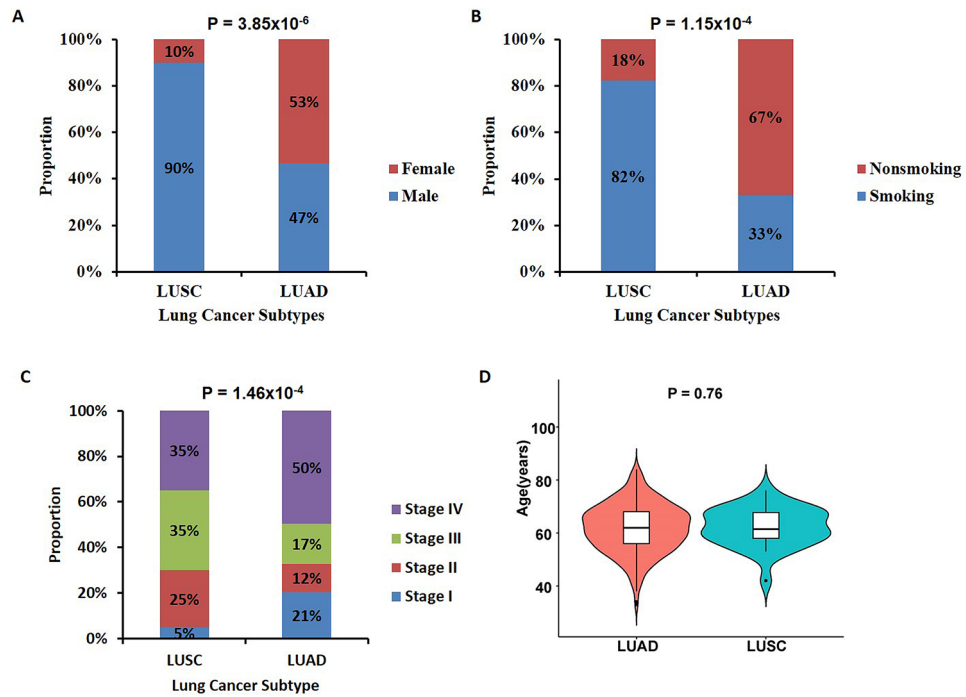


Figure 1. Association between tumor subtypes and clinical characteristics. (A) Association between tumor subtypes and sex; (B) Association between tumor subtypes and smoking status; (C) Correlation between tumor subtypes and tumor stage; (D) Correlation analysis between tumor subtypes and patient age. *LUAD* lung adenocarcinoma, *LUSC* lung squamous cell carcinoma, *LASC* lung adenosquamous cell carcinoma, *SCLC* small cell lung cancer.

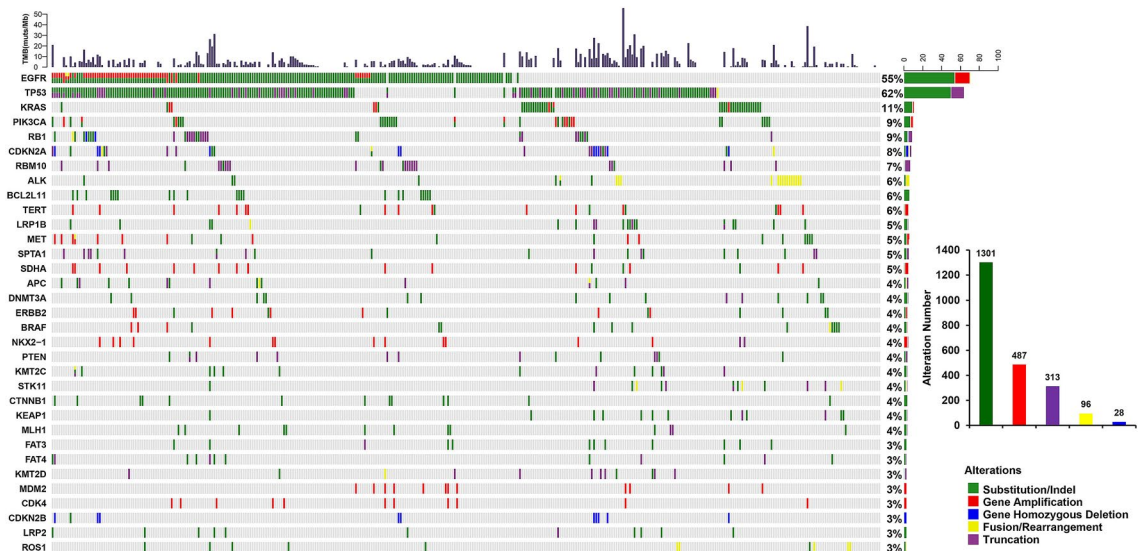


Figure 2. Mutational landscape of 371 Chinese lung cancer patients. The X-axis represents each patient tissue sample and the Y-axis represents each mutated gene. The bar graph above shows the tumor mutational burden (TMB) value of each sample, and the bar graph on the right shows the mutation frequency of each mutated gene. Statistical distribution of variation types is shown in the right column. Green represents substitution/indel, red represents gene amplification, blue represents gene homozygous deletion, yellow represents fusion/rearrangement, and purple represents truncation mutations.

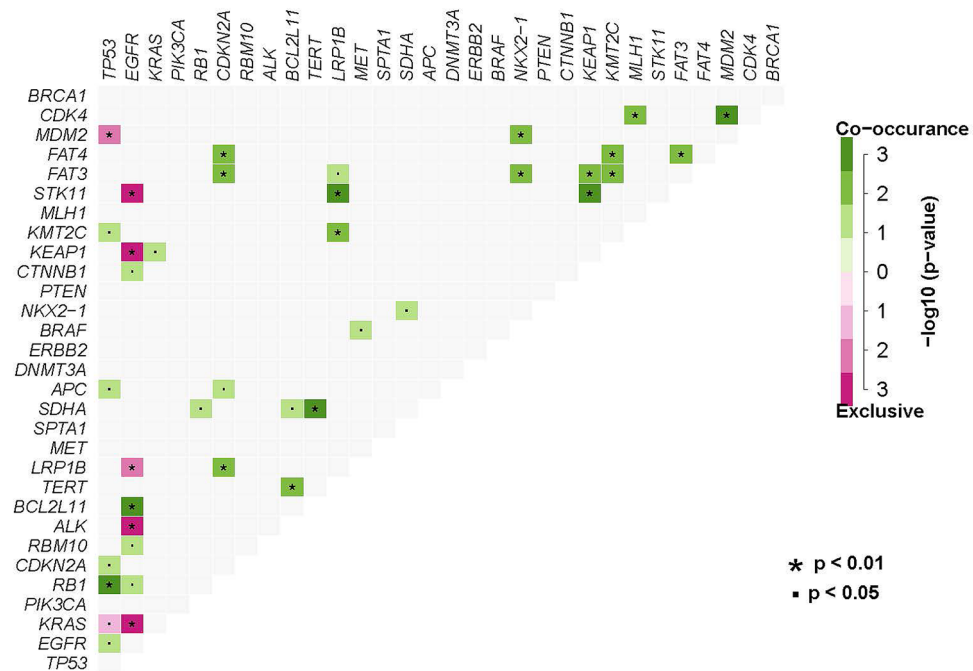


Figure 3. Co-occurring analysis of Chinese lung cancer patients. Green represents the co-occurrence mutations; pink represents the mutual exclusive mutations. $^{\bullet}P < 0.05$, $^{\ast}P < 0.01$.

with *TP53* mutations. In addition, the mutations in *BCL2L11*, *CTNNB1*, *RBM10*, and *RB1* co-occurred with *EGFR* mutations, while mutations in *STK11*, *KEAP1*, *LRP1B*, *ALK*, and *KRAS* were mutually exclusive with *EGFR* mutations. Notably, both *TP53* and *EGFR* mutations co-occurred with *RB1* mutations, and mutually exclusive with *KRAS* mutations (Fig. 3).

Differences between lung adenocarcinoma and lung squamous cell carcinoma. In this study, NSCLC represented nearly 90% of cases and mainly consisted of LUAD and LUSC. As shown in Fig. 4, there were many differences in the molecular characteristics of LUAD and LUSC. The most common mutations in LUAD and LUSC were *EGFR*, *TP53*, and *KRAS*, and *TP53*, *PIK3CA*, *CDKN2A*, *EGFR*, *CCND1*, *NFE2L2*, *FAM1358*, and *FGFR1*, respectively (Fig. 4A,B). In LUAD, the main mutation type of *PIK3CA* was SNV; in LUSC, it was mainly gene amplification. *ALK* fusion; *RBM10* truncation; *MET*, *TERT*, *NKX2-1*, *SDHA*, *CDK4*, and *MDM2* amplifications, and *CDKN2A* deletion were mainly identified in LUAD (Fig. 4A). *CCND1*, *FGFR1*, and *FGF3/4/13*, and *SOX2* amplifications were mainly identified in LUSC (Fig. 4B). Statistical analysis showed that the mutation frequency of *EGFR* was higher in LUAD than in LUSC ($P = 0.0015$), while the mutation frequencies of *TP53* ($P = 0.064$), *PIK3CA* ($P = 0.00014$), *NFE2L2* ($P = 0.0038$), *KMT2D* ($P = 0.0066$), *FGFR1* ($P = 0.023$), *CCND1* ($P = 0.033$), and *CDKN2A* ($P = 0.035$) were higher in LUSC than in LUAD (Fig. 5).

Age-related gene mutations in Chinese lung cancer patient. We examined the correlation between patient age and gene mutations. The results showed that patients with mutations in *ALK*, *ERBB2*, or *KMT2D* were younger than those without these mutations, while patients with mutations in *RBM10*, *NRAS*, *NF1*, *PIK3CA*, *MET*, *PBRM1*, *LRP2*, *NFE2L2*, or *CDKN2B* were older than those without these mutations. Statistical analysis showed that the mutation of these genes was significantly associated with the patient age (Fig. 6).

Correlations between mutated genes and smoking status, tumor stage, and sex in Chinese lung cancer patients. Based on the smoking status data, we analyzed the correlation between mutated genes and smoking status of patients. The most common mutated genes in smokers included *TP53*, *EGFR*, *KRAS*, *CDKN2A*, *LRP1B*, *ALK*, *BCL2L11*, *KEAP1*, *KMT2C*, *PIK3CA*, and *STK11*. The most common mutated genes in non-smokers were *EGFR*, *TP53*, *RB1*, *SDHA*, *RBM10*, and *TERT* (Table S2). *TP53* and *EGFR* mutations frequently occurred in both smokers and nonsmokers. Based on statistical analysis, the frequency of *EGFR* mutations was significantly higher in nonsmoking than smoking patients (Fig. 7A). The mutation frequencies of *CDKN2A*, *FAT1*, *FGFR1*, *NFE2L2*, *CCNE1*, *CCND1*, *SMARCA4*, *KEAP1*, *KMT2C*, and *STK11* were significantly higher in smokers than in nonsmokers (Fig. 7A).

According to the information on tumor stage, we combined 80 cases of stage I and II into a group, and 230 cases of stage III and IV into another group. Statistical analysis showed that *TP53* and *RB1* mutation frequencies were significantly higher in cases with tumor stages III and IV than in those with tumor stages I and II, while the mutation frequencies of *TAF1*, *LRP1B*, *SDHA*, *CBFB*, *BRIP1*, and *SMAD4* were significantly higher in cases with tumor stages I and II than in those with tumor stages III and IV (Fig. 7B).

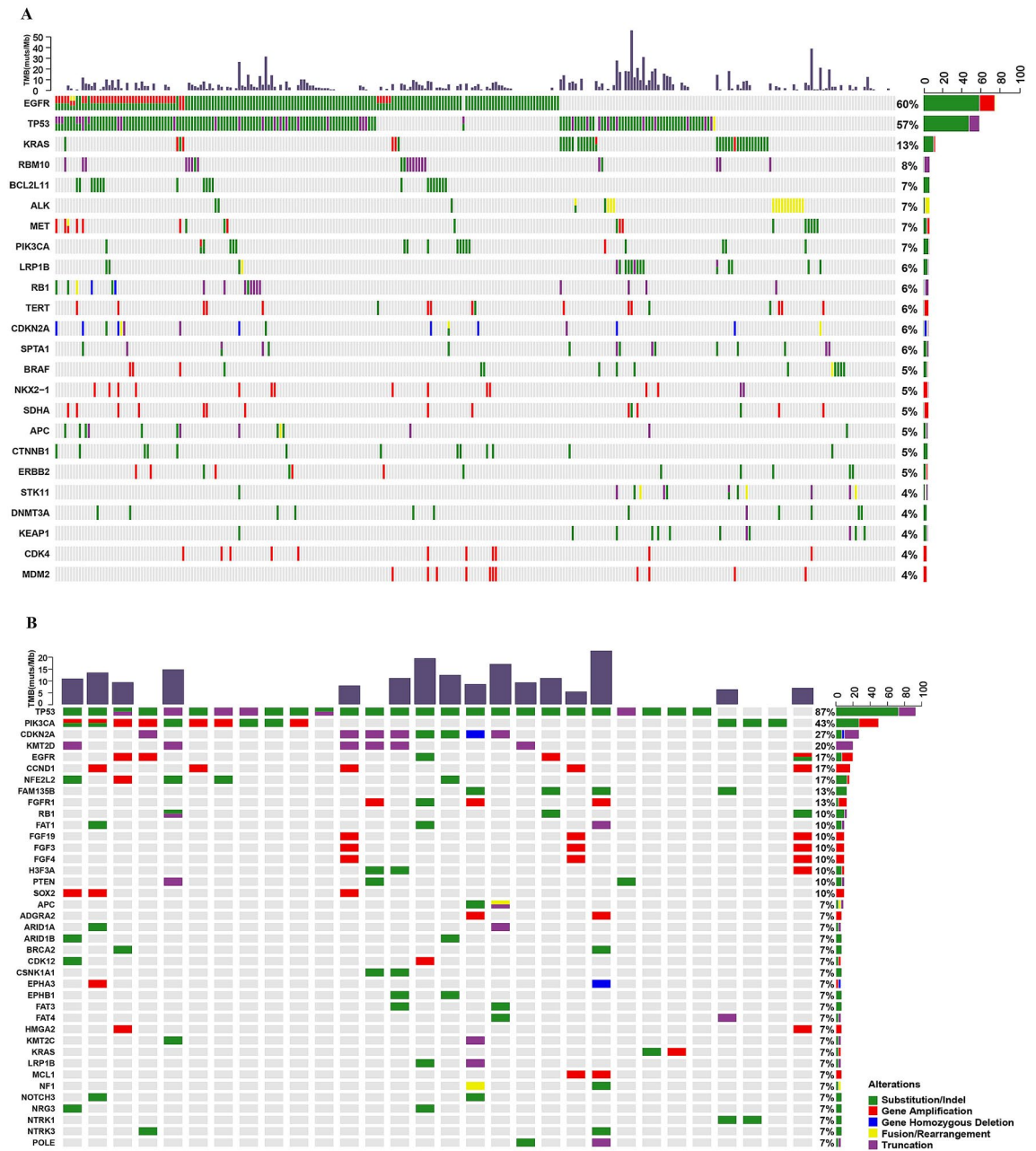


Figure 4. Mutational landscape of Chinese patients with lung adenocarcinoma (A) and lung squamous cell carcinoma (B). The X-axis represents each patient tissue sample and the Y-axis represents each mutated gene. The bar graph above shows the tumor mutational burden (TMB) value of each sample, and the bar graph on the right shows the mutation frequency of each mutated gene. Green represents substitution/Indel, red represents gene amplification, blue represents gene homozygous deletion, yellow represents fusion/rearrangement, and purple represents truncation mutations.

We also analyzed the association between mutated genes and sex. The statistical analysis results showed that EGFR mutation was significantly associated with sex (Fig. 7C).

Correlations between tumor mutation burden and clinical characteristics and mutated genes.

We measured the available TMB in 216 cases to explore the relationship between TMB and clinical characteristics, and TMB and clinically relevant mutations. The median TMB was 5.0 muts/Mb (range, 0–55.7 muts/Mb) (Table 1). We identified TMB-H in 49 cases (22.7%, 49/216) and TMB-L was identified in 167 cases (77.3%, 167/216). The median TMB in LUSC was higher than that in LUAD. Statistical analysis showed a significant association between TMB and tumor subtype (Fig. 8A). According to the age distribution of patients, we found that the TMB value increased gradually with the increase in age. Statistical analysis showed a positive correlation between age and TMB value in lung cancer (Fig. 8B). In this cohort, we also found that the median

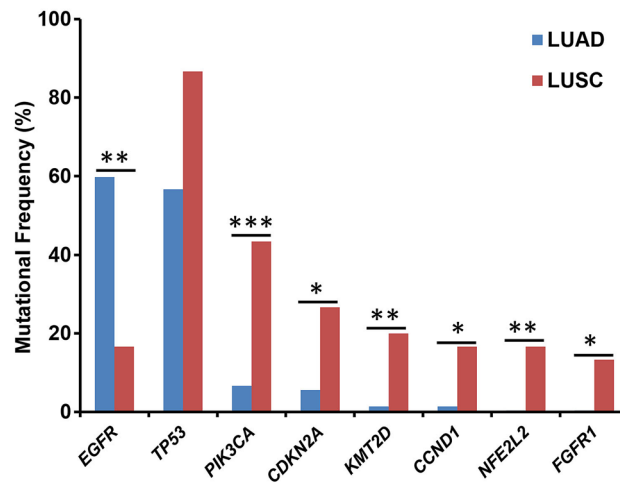


Figure 5. Correlation analysis of mutated genes and tumor subtype. The X-axis shows the mutated genes and the Y-axis represents the mutational frequency of each gene. LUAD and LUSC are represented by blue and red, respectively. * $P < 0.05$, ** $P < 0.01$, and *** $P < 0.001$. LUAD lung adenocarcinoma, LUSC lung squamous cell carcinoma.

TMB was higher in males than in females (7 mutations/Mb vs 4.3 mutations/Mb). The median TMB of smoking and non-smoking patients was 8.5 mutations/Mb and 3.8 mutations/Mb, respectively. Statistical analysis showed a significant association between TMB and sex and smoking status of patients (Fig. 8C,D).

Based on clinical relevance, we also analyzed TMB-related mutations. For each tested gene, patients were divided into mutant and wild type groups. Statistical analysis showed that mutations in *CDKN2A*, *LRP1B*, *LRP2*, *TP53*, and *EGFR* were significantly associated with TMB. Among these five genes, mutations in *CDKN2A*, *LRP1B*, *LRP2*, and *TP53* were associated with high TMB, while *EGFR* mutations were associated with low TMB (Fig. 9).

Characterization of *EGFR* mutations in patients resistant to icotinib/gefitinib. In this cohort, 203 patients were harbored *EGFR* mutations, with 77 patients receiving *EGFR*-TKIs treatment. Of the patients who received this treatment, we followed up 29 patients who treated with icotinib (375 mg/day) or gefitinib (250 mg/day). Among them, 22 patients developed disease progression within 6 months and were considered drug resistant, while 7 benefited from the therapy for more than 6 months and were considered drug sensitive. A total of 55 *EGFR* alterations were detected in these 29 patients, including 8 L858R, 16 T790M, 19 19del, and 12 *EGFR* amplifications. Among the drug-resistant patients, 12 had T790M mutation. The patients who did not have this mutation included 2 patients with L858R mutation (one of them harbored *ERBB2* amplification), 2 patients with *EGFR* amplification, 4 patients with 19del mutation (one of them harbored *ERBB2* amplification), and 2 patients with both 19del mutation and *EGFR* amplification. Among the drug-sensitive patients, 4 had T790M mutation and 3 did not have this mutation (2 carried 19del and 1 carried L858R mutation) (Table 2).

In addition to the T790M mutation, we found that the proportion of *EGFR* amplification in patients with drug resistance was higher than that in patients with drug sensitivity (40% vs 0%). We also analyzed mutations other than *EGFR* in the followed up patients. We found that *DNMT3A* and *NOTCH4* mutations were lower in the icotinib/gefitinib-resistant patients than those in the drug-sensitive patients (0% vs 28.6%, $P = 0.052$, for both) (Fig. 10).

Discussion

Lung cancer, which is multi-factorial and has various histological subtypes, is one of the most dangerous malignant tumors to human health and life. In recent years, the incidence and mortality rates of lung cancer have increased significantly in many countries^{1,30}, with the incidence in women increasing annually³¹. In addition, men are more likely to develop LUSC, while women are more likely to develop LUAD³². Regarding risk factors, smoking is one of the most important for lung cancer. Smoking has been reported to be significantly associated with LUSC³³. Our results also supported that smoking was significantly associated with LUSC and LASC.

A total of 371 lung cancer patients (187 males and 184 females) were included in this study. Most of them were LUAD patients, and the proportion of different sexes in this group was similar. However, in LUSC patients, the proportion of males was higher than females. This might be due to the high proportion of smokers among male patients. Moreover, the incidence rate of lung cancer increases with age³⁴. However, the median age of patients in this study was approximately 60 years and there was no significant difference in age distribution among different cancer subtypes, indicating that there was no correlation between tumor subtypes and patient age.

The continuous development of NGS technology facilitates the analysis of the landscape of cancer mutations. LUAD and LUSC are the two major subtypes of lung cancer and previous studies have shown that they have different molecular characteristics. The most common mutations in LUAD were *TP53*, *KRAS*, *KEAP1*, *STK11*, *EGFR*, *NF1*, and *BRAF*⁴; in LUSC, they were *TP53*, *MLL2* (*KMT2D*), *CDKN2A*, *PIK3CA*, *KEAP1* and *NFE2L2*⁵.

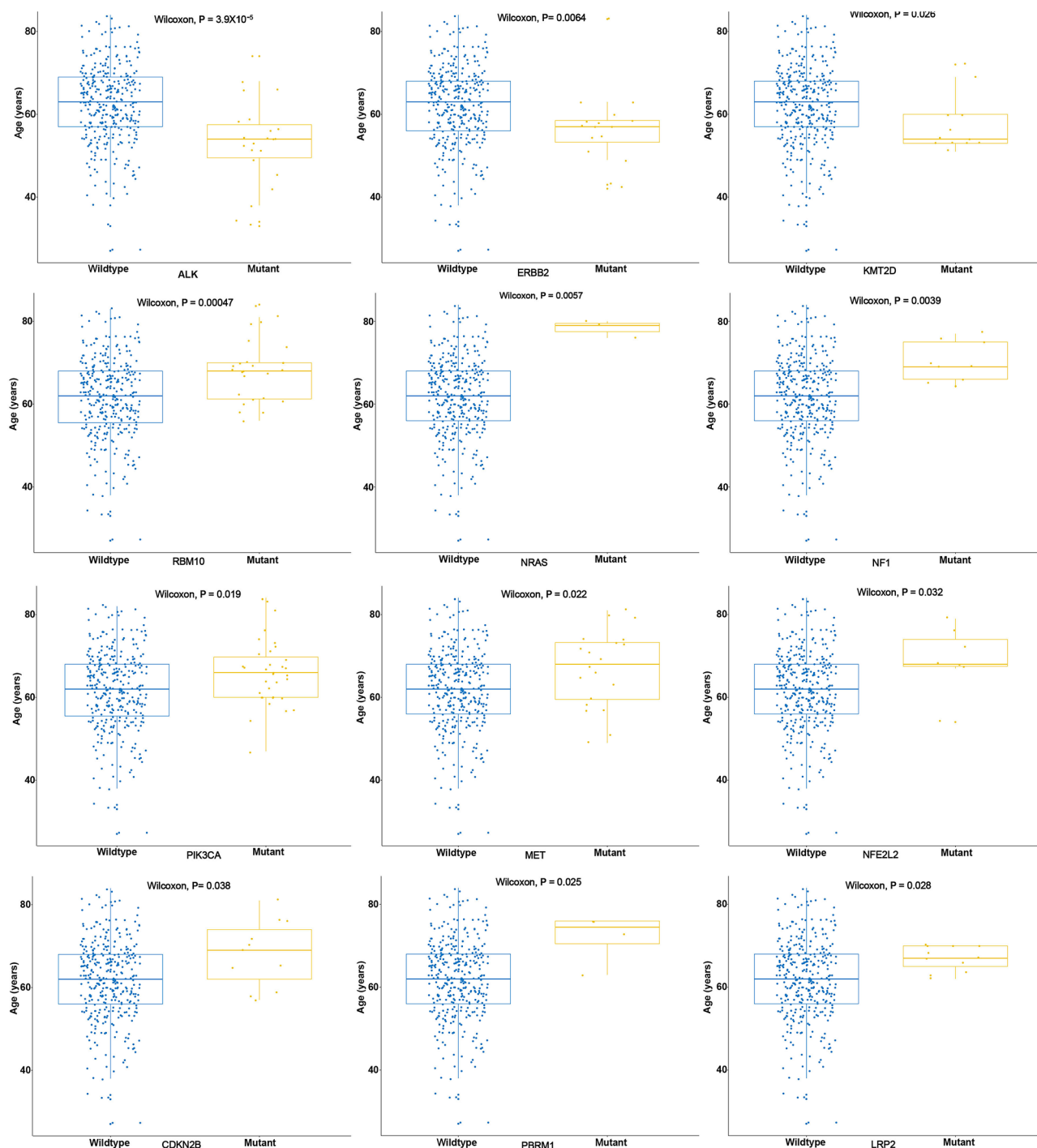


Figure 6. Correlation analysis of mutated genes and patients' age. The X-axis shows the mutated genes and the Y-axis represent the patients' age.

In contrast to the results of TCGA^{4,5}, we detected low mutational frequencies of *STK11*, *NF1*, and *BRAF* were detected in LUAD and low mutational frequency of *KEAP1* in both LUAD and LUSC, indicating the special molecular characteristics of Chinese lung cancer patients. In addition to observing fewer *EGFR* mutations in LUSC than in LUAD patients, which had been reported by Kim et al.³⁵, we also identified the different types of mutated genes in Chinese patients, such as *PIK3CA*, *ALK*, *RBM10*, *MET*, *TERT*, *NKX2-1*, *SDHA*, *CDK4*, and *MDM2* in LUAD; and *CCND1*, *FGFR1*, *FGF3/4/13*, and *SOX2* in LUSC. Recently, Ding et al. reported that mutations in *PIK3CA*, *FGFR1*, *CCND1*, and *CDKN2* mainly occurred in LUAD, while *TP53* mutations occurred in nearly 90% of LUSC patients⁸. Although we identified *TP53* mutations in 87% of LUSC patients, there was no significant difference in *TP53* mutational frequency between both types of cancer.

In lung cancer, *EGFR* mutations are frequently co-mutated with *TP53* and *RB1*, while *KRAS* mutations are frequently co-mutated with *STK11*, *KEAP1* and *RBM10*³⁶. Similarly, we found co-mutations of *EGFR*, *TP53*, and

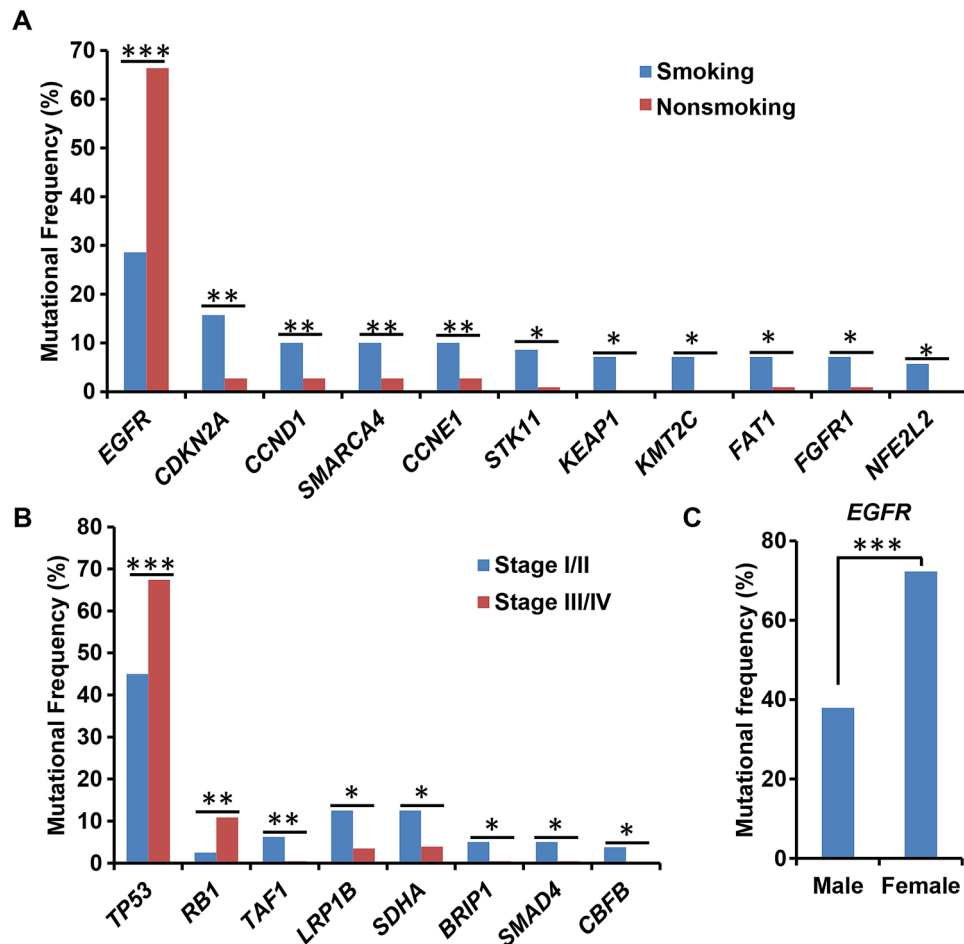


Figure 7. Correlation analysis of mutated genes and clinical characteristics. **(A)** Differences in mutated genes between smoking and nonsmoking patients. **(B)** Correlation between mutated genes and tumor stage. **(C)** Differences in mutational frequency of *EGFR* between male and female lung cancer patients. * $P < 0.05$, ** $P < 0.01$, and *** $P < 0.001$.

RB1, and the mutually exclusive mutations of *EGFR* with *STK11*, *KEAP1*, *ALK* and *KRAS* in this study. Concurrent *KRAS* mutations may lead to resistance to osimertinib and MEK inhibitor combined treatment³⁷. Mutual exclusive mutations of *EGFR* and *KRAS* may imply the potential opportunity to benefit from TKI-inhibitor therapy.

However, there was no co-mutation of *KRAS* with *STK11* and *KEAP1*. The inactivation of *TP53* and *RB1* is the molecular characterization of SCLC³⁸. In this study, 8 out of 11 SCLC patients harbored a co-mutation of *TP53* and *RB1*. All these results support the previous reported molecular features of lung cancer.

Furthermore, we found significantly different mutational frequencies of *NFE2L2* and *KMT2D*. *NFE2L2* is an important gene involved in the regulation of cell response to oxidative damage and chemotherapy⁵. A previous reports suggested that the *NFE2L2* mutation may be a biomarker for the special treatment of LUSC. Another study reported that the *KMT2D* mutation correlates with poor prognosis in NSCLC³⁹. In this way, the high frequency of *KMT2D* mutations indicated a poor prognosis of LUSC. However, the small number of LUSC samples is a limitation of this study and more expanded samples are needed to elucidate this association.

Age is an important factor for lung cancer and is often considered when selecting treatment⁴⁰. With the increasing proportion of young lung cancer patients⁴¹, more attention has been devoted to their diagnosis. Sacher et al. focused on the GAs of young lung cancer patients and identified that mutations in *EGFR*, *ALK*, and *ERBB2* trend to occur in younger NSCLC patients⁴². According to different age groups, Jiang et al. reported that mutations in *EGFR* and *TP53* were associated with age in Chinese NSCLC patients⁴³. In contrast to this study, we did not detect an association between age and *TP53* and *EGFR* mutations. However, we detected the correlation between age and *ALK* and *ERBB2* mutations, similar to the results of Sacher et al.⁴², which showed the reliability of our analysis. Furthermore, we identified associations between younger patients and *NRAS* and *KMT2D* mutations, and elderly patients and *RBM10*, *NF1*, *PIK3CA*, *MET*, *PBRM1*, *LRP2*, and *CDKN2B* mutations. These results contribute to the age-associated gene alteration data in lung cancer.

The mutational profile is different in smoking and nonsmoking patients. Mutations in *EGFR*, *ROS1*, and *ALK* mainly occur in nonsmoking patients, while mutations in *KRAS*, *TP53*, *BRAF*, *JAK2*, and *JAK3* mainly occur in

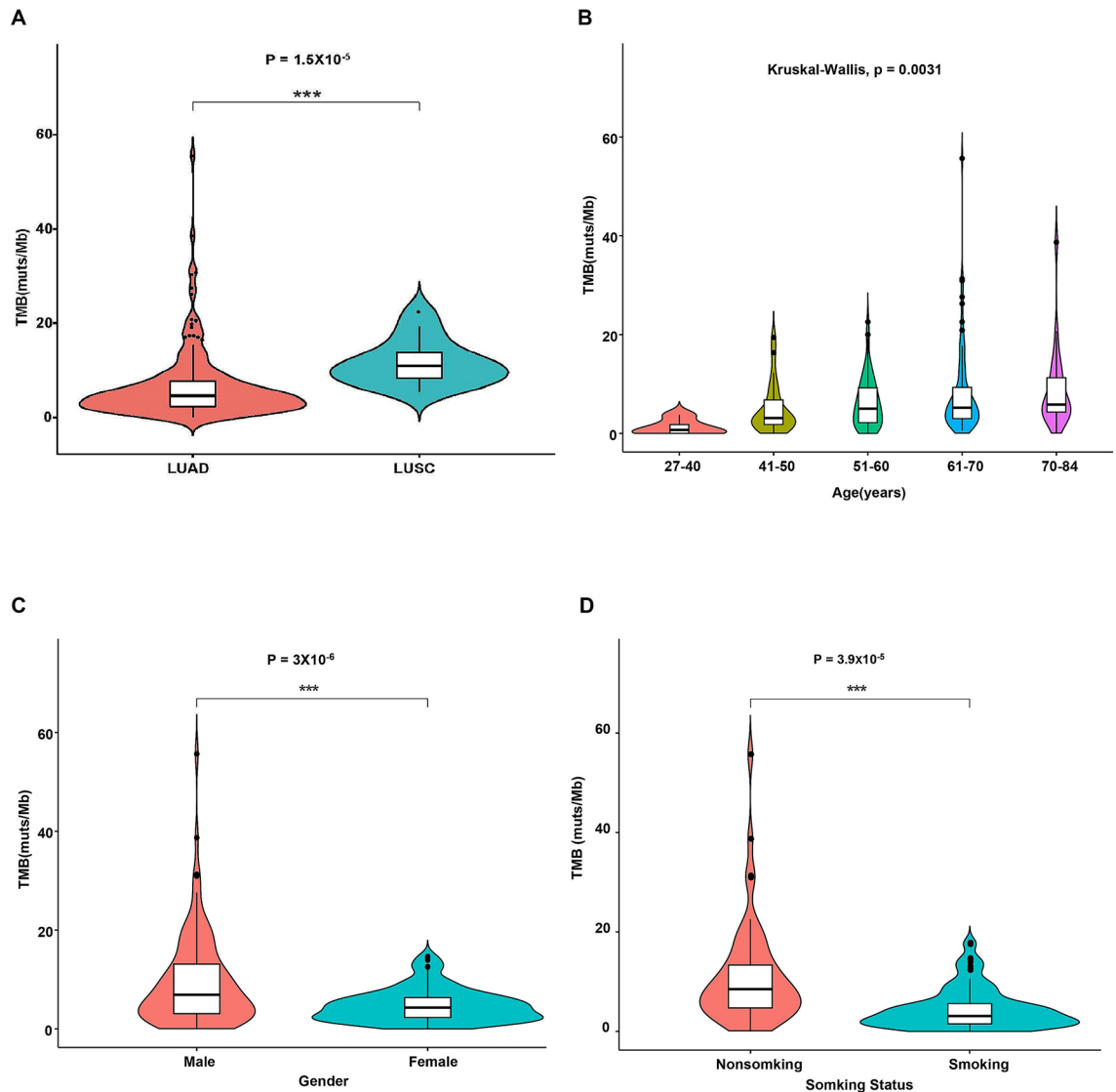


Figure 8. Association between TMB and clinical characteristics. (A) Association between TMB and tumor subtypes; (B) Association between TMB and age of patients; (C) Correlation analysis between TMB and sex; (D) Correlation analysis between TMB and smoking status. *LUAD* lung adenocarcinoma, *LUSC* lung squamous cell carcinoma, *LASC* lung adenosquamous cell carcinoma, *SCLC* small cell lung cancer, *TMB* tumor mutational burden.

smoking patients⁴⁴. However, only *EGFR* mutations were found to be associated with nonsmoking. Although the mutational frequencies of *KRAS* and *TP53* were also higher in smoking than those in nonsmoking patients in this study, the statistical analysis showed no significant difference (17.1% vs. 7.3%, $P = 0.052$; and 62.9% vs. 52.7%, $P = 0.22$, respectively). In contrast to previous studies^{43,44}, our results revealed a series of smoking-associated genes, including *CDKN2A*, *CCND1*, *SMARCA4*, *CCNE1*, *STK11*, *KEAP1*, *KMT2C*, *FAT1*, *FGFR1*, and *NFE2L2*. This discrepancy may be caused by regional differences among populations.

Previous studies have shown that female patients have a lower risk of cancer progression than male patients^{45,46}. Sex-related biomarkers could indicate specific treatment options. Similar to previous studies, we found that the mutational frequency of *EGFR* was significantly higher in female patients than that in male patients⁴. Tumor, lymph node and metastasis (TNM) staging are often used in treatment decisions and prognosis prediction of lung cancer patients⁴⁷. For NSCLC, stages I-II are considered early stages and are normally treated with surgery, while stages III-IV are advanced stages and are normally treated with concurrent chemoradiotherapy⁴⁸. Our results showed a correlation between *TP53* and *RB1* mutations and tumor stages III-IV. *TP53* and *RB1* are important regulators of cell cycle progression. *TP53* is the most common mutated gene in human cancers, and both *TP53* and *RB1* mutations are reported to be associated with poor prognosis of lung cancer patients^{49–52}. Our results indicate that these mutations may predict the prognosis of Chinese lung cancer patients.

Moreover, we showed a significant association between early tumor stage and mutations in *TAF1*, *LRP1B*, *SDHA*, *CBFB*, *BRIP1*, and *SMAD4*. Chen et al. reported that *LRP1B* mutation was associated with better survival in NSCLC patients treated with immunotherapy⁵³. Additionally, *SMAD4* expression is associated with

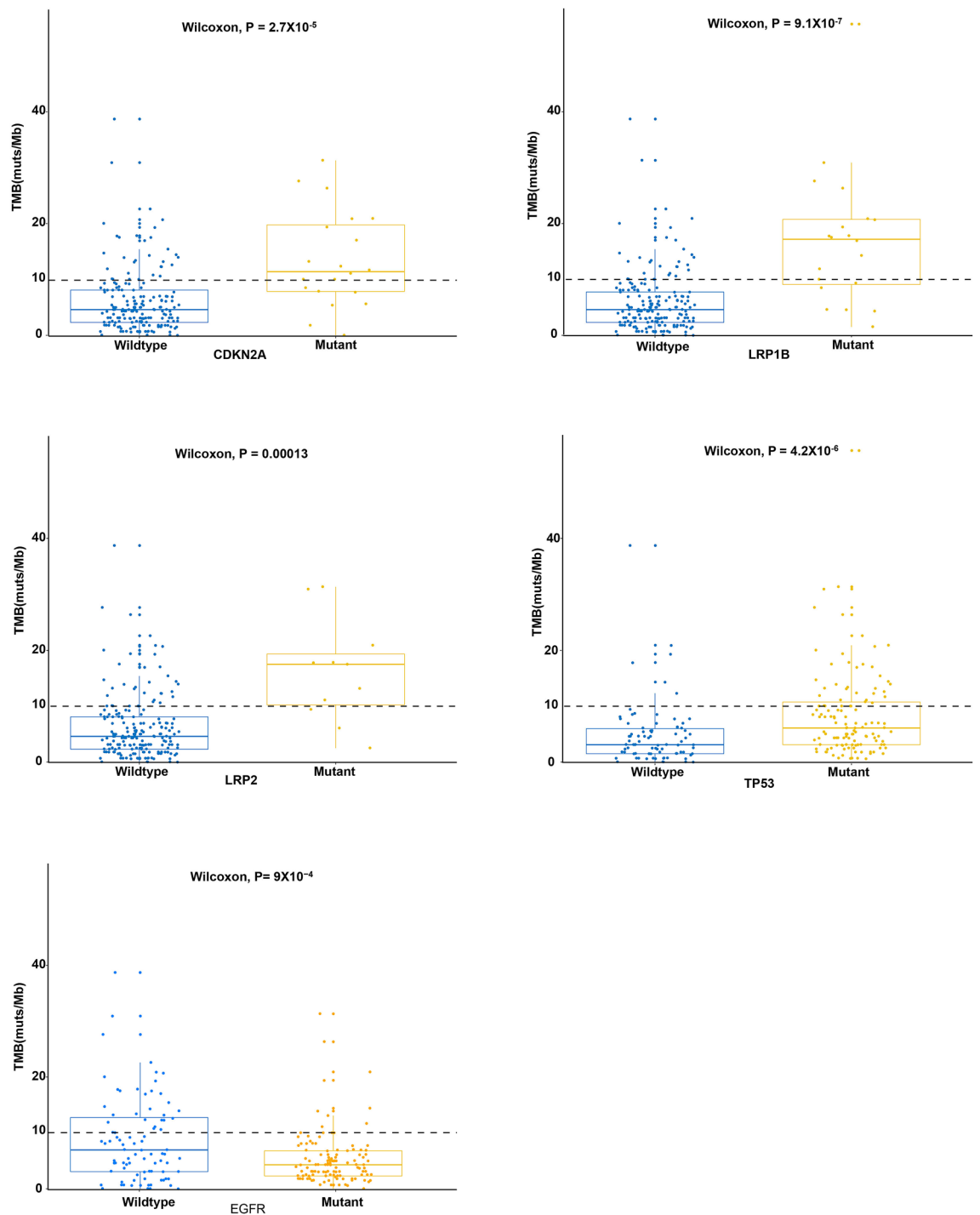


Figure 9. Correlation analysis between TMB and mutated genes. The X-axis shows the gene status and the Y-axis represents TMB values. TMB tumor mutational burden.

survival of patients with lung and pancreatic cancers⁵⁴, while *SDHA* is considered a tumor suppressor gene of paraganglioma⁵⁵. In this study, we first reported the correlation between *SDHA* mutation and tumor stage, indicating its potential predictive value. Although the correlation between *TAF1*, *BRIP1*, and *CBFβ* mutations and prognosis has been reported, only have been reported in lung cancer^{56–58}. Our results suggest that these genes may be related to prognosis in Chinese lung cancer patients. However, studies with a longer follow-up period are needed to elucidate this relationship.

TMB is a new biomarker that may further guide the selection of checkpoint inhibitors (CPI) for patients⁵⁹. A certain correlation between TMB and clinical characteristics has been reported. Wang et al. reported that the predictive power of TMB in lung cancer immunotherapy response was significantly better for women than for men⁶⁰. In addition, it has been reported that increased TMB is associated with increased age in many types of

Cases	Cancer subtype	Drug responses	Tumor stage	EGFR mutation			ERBB2 amplification	
1	LUAD	R	IV	T790M			L858R	
2	LUAD	R	III	T790M			L858R	
3	LUAD	R	IV				L858R	
4	LUAD	R	–	T790M			L858R	
5	LUAD	R	IV	T790M			L858R	
6	LUAD	R	III				L858R	yes
7	LUAD	R	IV	T790M	19 exon del	Amplification		
8	LUAD	R	–		19 exon del	Amplification		
9	LUAD	R	IV	T790M	19 exon del	Amplification		
10	SCLC	R	IV		19 exon del	Amplification		
11	LUAD	R	IV	T790M	19 exon del	Amplification		
12	LUAD	R	IV	T790M	19 exon del	Amplification		
13	LUAD	R	IV	T790M	19 exon del	Amplification		
14	LUSC	R	IV	T790M	19 exon del	Amplification		
15	LUAD	R	III		19 exon del		yes	
16	LUAD	R	IV	T790M	19 exon del			
17	LUAD	R	–		19 exon del			
18	LUAD	R	–		19 exon del			
19	NSCLC	R	IV		19 exon del			
20	LASC	R	IV			Amplification		
21	SCLC	R	IV			Amplification		
22	LUAD	R	–	T790M		Amplification	L858R	
23	LUAD	S	IV	T790M	19 exon del			
24	NSCLC	S	–	T790M	19 exon del			
25	LUAD	S	IV	T790M	19 exon del			
26	LUAD	S	IV		19 exon del			
27	LUAD	S	–		19 exon del			
28	LUAD	S	IV	T790M	19 exon del	Amplification		
29	LUAD	S	I				L858R	

Table 2. Lotinib/gefitinib response and EGFR mutations of 29 lung cancer patients. *LUAD* lung adenocarcinoma, *LUSC* lung squamous cell carcinoma, *LASC* lung adenosquamous cell carcinoma, *SCLC* small cell lung cancer.

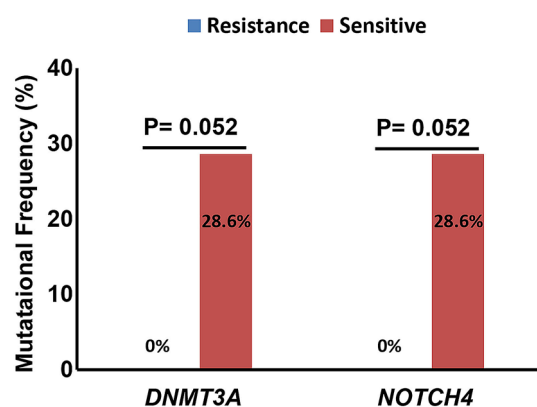


Figure 10. Specific DNMT3A and NOTCH4 mutations in icotinib/gefitinib resistant patients. The X-axis shows the mutated gene and the Y-axis represents mutational frequency of each gene in icotinib/gefitinib resistant and sensitive patients. *TMB* tumor mutational burden.

cancers⁶¹. Wang et al. reported associations between TMB and smoking history and age of patients with LUAD⁶². In NSCLC patients with TMB-H, non-smokers had a significantly better prognosis compared with smokers⁶³. However, there was no difference in TMB values between smoking and nonsmoking SCLC patients⁶⁴. Similar to a previous study, we identified associations between TMB and sex, age, and smoking status. Furthermore, our

results showed a significant association between TMB and tumor subtype. However, we also detected a correlation between smoking status and tumor subtype, indicating that the association between tumor subtype and TMB may be caused by the smoking status.

Moreover, TMB is associated with known DNA mismatch repair pathway genes (*MSH2*, *MSH6*, *MLH1*, and *PMS2*) and DNA polymerases (*POLE*)⁶¹. In this study, we failed to detect a correlation between TMB and these genes. However, statistical analysis showed the significant association between TMB and mutations in *EGFR*, *TP53*, *LRP1B*, *LRP2*, and *CDKN2A*, suggesting potential biomarkers for the prognosis of Chinese lung cancer patients. Particularly, *TP53* mutation status may be a useful biomarker for predicting the response to immunotherapy in different cancer types⁶⁵. Owada-Ozaki et al. reported that high TMB is associated with poor prognosis in NSCLC⁶⁶. These studies supported our results.

EGFR-mutated lung cancer is a special molecular subgroup of lung cancer in which most patients benefit from treatment with *EGFR*-TKIs⁶⁷. The clinical course of *EGFR* mutant lung cancer is significantly heterogeneous, and acquisition of *EGFR* T790M mutation is the most frequent reason for first- and second-generation *EGFR*-TKIs⁶⁸. The receptor tyrosine kinase or alternative downstream compounds activate survival tracks such as *MET* amplification, *ERBB2* amplification, and *IGF1R* activation, which are the main *EGFR*-independent reasons for *EGFR*-TKIs resistance^{69,70}.

Besides 19del and L858R, *EGFR* amplification is also frequently occurs in lung cancer. Recently, Chen et al. reported that an *EGFR*-amplified cervical squamous cell carcinoma patient benefited from afatinib therapy⁷¹. *EGFR* amplification was also reported to be associated with better OS, PFS, CR, and PR in LUAD patients treated with erlotinib⁷². In this study, 29 patients received the treatment of icotinib/gefitinib. In addition to the T790M mutation or *ERBB2* amplification, we found a higher proportion of *EGFR* amplification in *EGFR*-TKI-resistant patients than that in *EGFR*-TKI-sensitive patients. It is important to consider whether *EGFR* amplification is associated with the rapid development of icotinib/gefitinib resistance.

However, 6 patients with 19del or L858R mutations also rapidly developed *EGFR*-TKIs resistance. We found a high mutational proportion of *DNMT3a* and *NOTCH4* in these patients. *DNMT3a* plays an important role in methylation status. The Notch signaling pathway has an important regulatory role in a variety of tumor stem cells. Mutations in *DNMT3a* and *NOTCH4* have been reported to be associated with better prognosis in patients with LUAD and NSCLC, respectively^{73,74}. Similarly, the *EGFR*-TKI sensitivity of patients with these indicates a good prognosis. It is suggested that *DNMT3a* and *NOTCH4* mutations may be potential biomarkers to predict sensitivity to *EGFR*-TKIs.

In conclusion, we identified the comprehensive genomic features of 371 Chinese lung cancer patients and found that sex and smoking status were significantly associated with lung cancer subtype. Furthermore, we detected that certain gene mutations were associated with age, smoking status, tumor stage, and TMB value. We also suggested a series of biomarkers for potential therapy and prognosis, and indicated that *EGFR* amplification, *DNMT3a* mutation, and *NOTCH4* mutation may be used to predict *EGFR*-TKI resistance. Together, our research contributes to the comprehensive understanding of lung cancer molecular features and provides evidence for the developing and application of precise therapeutic strategies for Chinese lung cancer patients.

Data availability

The datasets used and analyzed in this study are available from the corresponding author upon reasonable request.

Received: 28 April 2020; Accepted: 19 October 2020

Published online: 20 November 2020

References

1. Ferlay, J. et al. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *Int. J. Cancer* **144**, 1941–1953. <https://doi.org/10.1002/ijc.31937> (2019).
2. Beadsmoore, C. J. & Screaton, N. J. Classification, staging and prognosis of lung cancer. *Eur. J. Radiol.* **45**, 8–17. [https://doi.org/10.1016/s0720-048x\(02\)00287-5](https://doi.org/10.1016/s0720-048x(02)00287-5) (2003).
3. Biring, S. S. & Peake, M. D. Symptoms and the early diagnosis of lung cancer. *Thorax* **60**, 268–269. <https://doi.org/10.1136/thx.2004.032698> (2005).
4. The Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**, 543–550. <https://doi.org/10.1038/nature13385> (2014).
5. The Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489**, 519–525. <https://doi.org/10.1038/nature11404> (2012).
6. Saito, M., Suzuki, H., Kono, K., Takenoshita, S. & Kohno, T. Treatment of lung adenocarcinoma by molecular-targeted therapy and immunotherapy. *Surg. Today* **48**, 1–8. <https://doi.org/10.1007/s00595-017-1497-7> (2018).
7. Pilotto, S. et al. Integrating the molecular background of targeted therapy and immunotherapy in lung cancer: a way to explore the impact of mutational landscape on tumor immunogenicity. *Transl. Lung Cancer Res.* **4**, 721–727. <https://doi.org/10.3978/j.issn.2218-6751.2015.10.11> (2015).
8. Ding, Y. et al. Comparative study on the mutational profile of adenocarcinoma and squamous cell carcinoma predominant histologic subtypes in Chinese non-small cell lung cancer patients. *Thorac. Cancer* <https://doi.org/10.1111/1759-7714.13208> (2019).
9. Mok, T. S. Personalized medicine in lung cancer: what we need to know. *Nat. Rev. Clin. Oncol.* **8**, 661–668. <https://doi.org/10.1038/nrclinonc.2011.126> (2011).
10. Kohno, T. et al. KIF5B-RET fusions in lung adenocarcinoma. *Nat. Med.* **18**, 375–377. <https://doi.org/10.1038/nm.2644> (2012).
11. Cai, W. et al. KIF5B-RET fusions in Chinese patients with non-small cell lung cancer. *Cancer* **119**, 1486–1494. <https://doi.org/10.1002/ncr.27940> (2013).
12. Lee, J. J. et al. Tracing oncogene rearrangements in the mutational history of lung adenocarcinoma. *Cell* **177**, 1842–1857.e1821. <https://doi.org/10.1016/j.cell.2019.05.013> (2019).
13. Seo, J. S. et al. The transcriptional landscape and mutational profile of lung adenocarcinoma. *Genome Res.* **22**, 2109–2119. <https://doi.org/10.1101/gr.145144.112> (2012).

14. Chapman, A. M., Sun, K. Y., Ruestow, P., Cowan, D. M. & Madl, A. K. Lung cancer mutation profile of EGFR, ALK, and KRAS: meta-analysis and comparison of never and ever smokers. *Lung Cancer* **102**, 122–134. <https://doi.org/10.1016/j.lungcan.2016.10.010> (2016).
15. Vaishnavi, A. *et al.* Oncogenic and drug-sensitive NTRK1 rearrangements in lung cancer. *Nat. Med.* **19**, 1469–1472. <https://doi.org/10.1038/nm.3352> (2013).
16. Suzuki, A. *et al.* Identification and characterization of cancer mutations in Japanese lung adenocarcinoma without sequencing of normal tissue counterparts. *PLoS ONE* **8**, e73484. <https://doi.org/10.1371/journal.pone.0073484> (2013).
17. Colombino, M. *et al.* EGFR, KRAS, BRAF, ALK, and cMET genetic alterations in 1440 Sardinian patients with lung adenocarcinoma. *BMC Pulm. Med.* **19**, 209. <https://doi.org/10.1186/s12890-019-0964-x> (2019).
18. Bittner, N., Ostoros, G. & Geczi, L. New treatment options for lung adenocarcinoma—in view of molecular background. *Pathol. Oncol. Res.* **20**, 11–25. <https://doi.org/10.1007/s12253-013-9719-9> (2014).
19. Mok, T. S. *et al.* Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. *N. Engl. J. Med.* **361**, 947–957. <https://doi.org/10.1056/NEJMoa0810699> (2009).
20. Sequist, L. V. *et al.* Phase III study of afatinib or cisplatin plus pemetrexed in patients with metastatic lung adenocarcinoma with EGFR mutations. *J. Clin. Oncol.* **31**, 3327–3334. <https://doi.org/10.1200/jco.2012.44.2806> (2013).
21. Colabufo, N. A. *et al.* EGFR tyrosine kinase inhibitors and multidrug resistance: perspectives. *Front. Biosci.* **16**, 1811–1823. <https://doi.org/10.2741/3823> (2011).
22. Janne, P. A. *et al.* AZD9291 in EGFR inhibitor-resistant non-small-cell lung cancer. *N. Engl. J. Med.* **372**, 1689–1699. <https://doi.org/10.1056/NEJMoa1411817> (2015).
23. Mok, T. S. *et al.* Osimertinib or platinum-pemetrexed in EGFR T790M-positive lung cancer. *N. Engl. J. Med.* **376**, 629–640. <https://doi.org/10.1056/NEJMoa1612674> (2017).
24. Fernandez-Cuesta, L. *et al.* CD74-NRG1 fusions in lung adenocarcinoma. *Cancer Discov.* **4**, 415–422. <https://doi.org/10.1158/2159-8290.cd-13-0633> (2014).
25. Paynter, R. A. *et al.* Accuracy of multiplexed Illumina platform-based single-nucleotide polymorphism genotyping compared between genomic and whole genome amplified DNA collected from multiple sources. *Cancer Epidemiol. Biomark. Prev.* **15**, 2533–2536. <https://doi.org/10.1158/1055-9965.epi-06-0219> (2006).
26. Press, O. A. *et al.* Gender-related survival differences associated with EGFR polymorphisms in metastatic colon cancer. *Can. Res.* **68**, 3037–3042. <https://doi.org/10.1158/0008-5472.can-07-2718> (2008).
27. Frampton, G. M. *et al.* Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. *Nat. Biotechnol.* **31**, 1023–1031. <https://doi.org/10.1038/nbt.2696> (2013).
28. Cao, J. *et al.* An accurate and comprehensive clinical sequencing assay for cancer targeted and immunotherapies. *Oncologist* **24**, e1294–e1302. <https://doi.org/10.1634/theoncologist.2019-0236> (2019).
29. Detterbeck, F. C., Boffa, D. J., Kim, A. W. & Tanoue, L. T. The Eighth edition lung cancer stage classification. *Chest* **151**, 193–203. <https://doi.org/10.1016/j.chest.2016.10.010> (2017).
30. Gupta, S. *et al.* International trends in the incidence of cancer among adolescents and young adults. *J. Natl. Cancer Inst.* <https://doi.org/10.1093/jnci/djaa007> (2020).
31. Chen, K. *et al.* Twenty-year secular changes in sex specific lung cancer incidence rates in an urban Chinese population. *Lung Cancer* **51**, 13–19. <https://doi.org/10.1016/j.lungcan.2005.08.013> (2006).
32. Cerfolio, R. J. *et al.* Women with pathologic stage I, II, and III non-small cell lung cancer have better survival than men. *Chest* **130**, 1796–1802. <https://doi.org/10.1378/chest.130.6.1796> (2006).
33. Egawa, H. *et al.* Radiation and smoking effects on lung cancer incidence by histological types among atomic bomb survivors. *Radiat. Res.* **178**, 191–201. <https://doi.org/10.1667/rr2819.1> (2012).
34. Torre, L. A., Siegel, R. L. & Jemal, A. Lung cancer statistics. *Adv. Exp. Med. Biol.* **893**, 1–19. https://doi.org/10.1007/978-3-319-24223-1_1 (2016).
35. Kim, Y. *et al.* Integrative and comparative genomic analysis of lung squamous cell carcinomas in East Asian patients. *J. Clin. Oncol.* **32**, 121–128. <https://doi.org/10.1200/jco.2013.50.8556> (2014).
36. Skoulidis, F. & Heymach, J. V. Co-occurring genomic alterations in non-small-cell lung cancer biology and therapy. *Nat. Rev. Cancer* **19**, 495–509. <https://doi.org/10.1038/s41568-019-0179-8> (2019).
37. Del Re, M. *et al.* Implications of KRAS mutations in acquired resistance to treatment in NSCLC. *Oncotarget* **9**, 6630–6643. <https://doi.org/10.18632/oncotarget.23553> (2017).
38. Gazdar, A. F., Bunn, P. A. & Minna, J. D. Small-cell lung cancer: what we know, what we need to know and the path forward. *Nat. Rev. Cancer* **17**, 725–737. <https://doi.org/10.1038/nrc.2017.87> (2017).
39. Ardeshir-Larijani, F. *et al.* KMT2D mutation is associated with poor prognosis in non-small-cell lung cancer. *Clin. Lung Cancer* **19**, e489–e501. <https://doi.org/10.1016/j.clc.2018.03.005> (2018).
40. Jorgensen, T. L., Hallas, J., Friis, S. & Herrstedt, J. Comorbidity in elderly cancer patients in relation to overall and cancer-specific mortality. *Br. J. Cancer* **106**, 1353–1360. <https://doi.org/10.1038/bjc.2012.46> (2012).
41. Alonso, R. *et al.* Lung cancer incidence trends in Uruguay 1990–2014: an age-period-cohort analysis. *Cancer Epidemiol.* **55**, 17–22. <https://doi.org/10.1016/j.canep.2018.04.012> (2018).
42. Sacher, A. G. *et al.* Association between younger age and targetable genomic alterations and prognosis in non-small-cell lung cancer. *JAMA Oncol.* **2**, 313–320. <https://doi.org/10.1001/jamaoncol.2015.4482> (2016).
43. Jiang, R. *et al.* Validating a targeted next-generation sequencing assay and profiling somatic variants in Chinese non-small cell lung cancer patients. *Sci. Rep.* **10**, 2070. <https://doi.org/10.1038/s41598-020-58819-5> (2020).
44. Govindan, R. *et al.* Genomic landscape of non-small cell lung cancer in smokers and never-smokers. *Cell* **150**, 1121–1134. <https://doi.org/10.1016/j.cell.2012.08.024> (2012).
45. Hsu, L. H. *et al.* Sex-associated differences in non-small cell lung cancer in the new era: is gender an independent prognostic factor? *Lung Cancer* **66**, 262–267. <https://doi.org/10.1016/j.lungcan.2009.01.020> (2009).
46. Pinto, J. A. *et al.* Gender and outcomes in non-small cell lung cancer: an old prognostic variable comes back for targeted therapy and immunotherapy?. *ESMO Open* **3**, e000344. <https://doi.org/10.1136/esmoopen-2018-000344> (2018).
47. Paesmans, M. *et al.* Prognostic factors for survival in advanced non-small-cell lung cancer: univariate and multivariate analyses including recursive partitioning and amalgamation algorithms in 1,052 patients The European Lung Cancer Working Party. *J. Clin. Oncol.* **13**, 1221–1230. <https://doi.org/10.1200/jco.1995.13.5.1221> (1995).
48. Choi, H. S. *et al.* Application of the new 8th TNM staging system for non-small cell lung cancer: treated with curative concurrent chemoradiotherapy. *Radiat. Oncol.* **12**, 122. <https://doi.org/10.1186/s13014-017-0848-2> (2017).
49. Viktorsson, K., De Petris, L. & Lewensohn, R. The role of p53 in treatment responses of lung cancer. *Biochem. Biophys. Res. Commun.* **331**, 868–880. <https://doi.org/10.1016/j.bbrc.2005.03.192> (2005).
50. Ma, X. *et al.* Prognostic and predictive effect of TP53 mutations in patients with non-small cell lung cancer from adjuvant cisplatin-based therapy randomized trials: A LACE-bio pooled analysis. *J. Thor. Oncol.* **11**, 850–861. <https://doi.org/10.1016/j.jtho.2016.02.002> (2016).
51. Burke, L. *et al.* Prognostic implications of molecular and immunohistochemical profiles of the Rb and p53 cell cycle regulatory pathways in primary non-small cell lung carcinoma. *Clin. Cancer Res.* **11**, 232–241 (2005).

52. Kim, Y. *et al.* Concurrent genetic alterations predict the progression to target therapy in EGFR-mutated advanced NSCLC. *J. Thor. Oncol.* **14**, 193–202. <https://doi.org/10.1016/j.jtho.2018.10.150> (2019).
53. Chen, H. *et al.* Association of LRP1B mutation with tumor mutation burden and outcomes in melanoma and non-small cell lung cancer patients treated with immune check-point blockades. *Front. Immunol.* **10**, 1113. <https://doi.org/10.3389/fimmu.2019.01113> (2019).
54. Liu, J. *et al.* ErbB2 pathway activation upon Smad4 loss promotes lung tumor growth and metastasis. *Cell Rep.* **10**, 1599–1613. <https://doi.org/10.1016/j.celrep.2015.02.014> (2015).
55. Burnichon, N. *et al.* SDHA is a tumor suppressor gene causing paraganglioma. *Hum. Mol. Genet.* **19**, 3011–3020. <https://doi.org/10.1093/hmg/ddq206> (2010).
56. Cai, C. *et al.* Identification of TAF1, SAT1, and ARHGGEF9 as DNA methylation biomarkers for hepatocellular carcinoma. *J. Cell. Physiol.* **235**, 611–618. <https://doi.org/10.1002/jcp.28999> (2020).
57. Pooley, K. A. *et al.* Common single-nucleotide polymorphisms in DNA double-strand break repair genes and breast cancer risk. *Cancer Epidemiol. Biomark. Prev.* **17**, 3482–3489. <https://doi.org/10.1158/1055-9965.epi-08-0594> (2008).
58. Andersen, C. L. *et al.* Dysregulation of the transcription factors SOX4, CBFβ and SMARCC1 correlates with outcome of colorectal cancer. *Br. J. Cancer* **100**, 511–523. <https://doi.org/10.1038/sj.bjc.6604884> (2009).
59. Riaz, N. *et al.* Tumor and microenvironment evolution during immunotherapy with nivolumab. *Cell* **171**, 934–949.e916. <https://doi.org/10.1016/j.cell.2017.09.028> (2017).
60. Wang, S., Zhang, J., He, Z., Wu, K. & Liu, X. S. The predictive power of tumor mutational burden in lung cancer immunotherapy response is influenced by patients' sex. *Int. J. Cancer* **145**, 2840–2849. <https://doi.org/10.1002/ijc.32327> (2019).
61. Chalmers, Z. R. *et al.* Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome medicine* **9**, 34. <https://doi.org/10.1186/s13073-017-0424-2> (2017).
62. Wang, C. *et al.* Molecular subtyping and prognostic assessment based on tumor mutation burden in patients with lung adenocarcinomas. *Int. J. Mol. Sci.* **20**, 4251. <https://doi.org/10.3390/ijms20174251> (2019).
63. Sardari Nia, P. *et al.* Prognostic value of smoking status in operated non-small cell lung cancer. *Lung Cancer* **47**, 351–359. <https://doi.org/10.1016/j.lungcan.2004.08.011> (2005).
64. Su, S. *et al.* Tumor mutational burden and genomic alterations in chinese small cell lung cancer measured by whole-exome sequencing. *Biomed. Res. Int.* **6096350–6096350**, 2019. <https://doi.org/10.1155/2019/6096350> (2019).
65. Li, L., Li, M. & Wang, X. Cancer type-dependent correlations between TP53 mutations and antitumor immunity. *DNA Repair* **88**, 102785. <https://doi.org/10.1016/j.dnarep.2020.102785> (2020).
66. Owada-Ozaki, Y. *et al.* prognostic impact of tumor mutation burden in patients with completely resected non-small cell lung cancer: brief report. *J. Thor. Oncol.* **13**, 1217–1221. <https://doi.org/10.1016/j.jtho.2018.04.003> (2018).
67. Rosell, R. *et al.* Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. *Lancet Oncol.* **13**, 239–246. [https://doi.org/10.1016/s1470-2045\(11\)70393-x](https://doi.org/10.1016/s1470-2045(11)70393-x) (2012).
68. Yu, H. A. *et al.* Concurrent alterations in EGFR-mutant lung cancers associated with resistance to EGFR kinase inhibitors and characterization of MTOR as a mediator of resistance. *Clin. Cancer Res.* **24**, 3108–3118. <https://doi.org/10.1158/1078-0432.ccr-17-2961> (2018).
69. Niederst, M. J. & Engelman, J. A. Bypass mechanisms of resistance to receptor tyrosine kinase inhibition in lung cancer. *Sci. Signal.* **6**, 6. <https://doi.org/10.1126/scisignal.2004652> (2013).
70. Liu, Q. *et al.* EGFR-TKIs resistance via EGFR-independent signaling pathways. *Mol. Cancer* **17**, 53. <https://doi.org/10.1186/s12943-018-0793-1> (2018).
71. Chen, Q. *et al.* An EGFR-amplified cervical squamous cell carcinoma patient with pulmonary metastasis benefits from afatinib: a case report. *OncoTargets Ther.* **13**, 1845–1849. <https://doi.org/10.2147/ott.s236382> (2020).
72. Ruiz-Patino, A. *et al.* EGFR amplification and sensitizing mutations correlate with survival in lung adenocarcinoma patients treated with erlotinib (MutP-CLICaP). *Target. Oncol.* **13**, 621–629. <https://doi.org/10.1007/s11523-018-0594-x> (2018).
73. Husni, R. E. *et al.* DNMT3a expression pattern and its prognostic value in lung adenocarcinoma. *Lung Cancer* **97**, 59–65. <https://doi.org/10.1016/j.lungcan.2016.04.018> (2016).
74. Xu, Q. *et al.* Association between single nucleotide polymorphisms of NOTCH signaling pathway-related genes and the prognosis of NSCLC. *Cancer Manage. Res.* **11**, 6895–6905. <https://doi.org/10.2147/cmar.s197747> (2019).

Acknowledgements

The authors are grateful to all the study participants, patients, and their family members for their contributions and support.

Author contributions

X.L., W.L., R.H., G.F., X.H., J.Z., K.W., Z.D., Y.Z., and A.Z. collected patient consents and samples and analyzed data; X.L., W.L., X.S., S.Y. and L.Z. contributed to bioinformatics analysis and wrote the manuscript; Y.S. designed and supervised the study. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-76791-y>.

Correspondence and requests for materials should be addressed to Y.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020