



## Comprehensive profiling of EGFR mutation subtypes reveals genomic-clinical associations in non-small-cell lung cancer patients on first-generation EGFR inhibitors

Yongkang Bai<sup>a,1</sup>, Xiang Liu<sup>b,1</sup>, Limin Zheng<sup>c,1</sup>, Song Wang<sup>d</sup>, Junli Zhang<sup>d</sup>, Shi Xiong<sup>a</sup>, Pengfei Zhang<sup>a</sup>, Zichen Jiao<sup>a</sup>, Gefei Zhao<sup>a</sup>, Chu Zhou<sup>a</sup>, Jiaohui Pang<sup>d</sup>, Yang Xu<sup>d</sup>, Qiuxiang Ou<sup>d</sup>, Yuan Mao<sup>e,f,\*</sup>, Louqian Zhang<sup>a,\*\*</sup>

<sup>a</sup> Department of Thoracic Surgery, Nanjing Drum Tower Hospital, Medical School, Nanjing University, Nanjing, Jiangsu, China

<sup>b</sup> Nanjing Drum Tower Hospital Group Suqian Hospital, Nanjing, Jiangsu, China

<sup>c</sup> Department of Clinical Pharmacy, China Pharmaceutical University, Nanjing, Jiangsu, China

<sup>d</sup> Geneseeq Research Institute, Nanjing Geneseeq Technology Inc., Nanjing, Jiangsu, China

<sup>e</sup> Department of Oncology, The Fourth Affiliated Hospital of Nanjing Medical University, Nanjing, Jiangsu, China

<sup>f</sup> Department of Oncology, Geriatric Hospital of Nanjing Medical University, Nanjing, Jiangsu, China

### ARTICLE INFO

#### Keywords:

Common EGFR  
Uncommon EGFR  
NSCLC  
EGFR-TKI

### ABSTRACT

Common sensitizing mutations in epidermal growth factor receptor (cEGFR), including exon 19 deletions (19-Del) and exon 21 L858R substitution, are associated with high sensitivity to EGFR-TKIs in NSCLC patients. The treatment for NSCLC patients with uncommon EGFR (uEGFR) mutations remains a subject of debate due to heterogeneity in treatment responses. In this manuscript, the targeted next-generation sequencing (NGS) data of a large cohort of EGFR-mutated NSCLC patients was assessed to elucidate genomic profiles of tumors carrying cEGFR or uEGFR mutations. The results showed that NSCLC patients with uEGFR mutations were more likely to harbor co-occurring genetic alterations in the Hippo pathway and a higher TMB compared with cEGFR-positive patients. Smoking-related mutations were found to significantly enriched in uEGFR-positive patients. Subgroup analyses were performed to identify potential prognostic biomarkers in patients harboring various EGFR subtype mutations. L858R-positive patients with co-existing ARID2 mutations had shorter progression-free survival (PFS) than those who were L858R- or 19-Del-positive but ARID2-negative (median: 2.3 vs. 12.0 vs. 8.0 months,  $P = 0.038$ ). Furthermore, mutational profiles, such as top frequently mutated genes and mutational signatures of patients with various EGFR subtype mutations were significantly different. Our study analyzed the mutational landscape of NSCLC patients harboring cEGFR and uEGFR mutations, revealing specific genomic characteristics associated with uEGFR mutations that might explain the poor prognosis of first-generation EGFR-TKIs.

### Introduction

Lung cancer is the most common cause of cancer-related mortality worldwide, with non-small cell lung cancer (NSCLC) accounting for ~80% of newly diagnosed lung cancer cases annually [1–3]. The epidermal growth factor receptor tyrosine kinase inhibitors (EGFR-TKIs)

have demonstrated profound clinical efficacies in NSCLC patients harboring EGFR exon 19 deletion (19-Del) or L858R mutations, showing superior survival benefit over platinum-based chemotherapy and fewer side effects [4,5].

Activation of EGFR can promote cell proliferation, angiogenesis, tumor invasion, and metastasis, therefore playing an essential role in the

**Abbreviations:** EGFR, epidermal growth factor receptor; cEGFR, common EGFR; uEGFR, uncommon EGFR; 19-Del, EGFR exon 19 deletion; 20-Ins, EGFR exon 20 insertions; TKI, tyrosine kinase inhibitor; NSCLC, non-small-cell lung cancer; NGS, next-generation sequencing; TMB, tumor mutational burden; FFPE, formalin-fixed paraffin-embedded; SNVs, single-nucleotide variations; MAF, mutant allele frequency; COSMIC, Catalog of Somatic Mutations in Cancer; PFS, progression-free survival; OS, overall survival.

\* Corresponding author at: Department of Oncology, The Fourth Affiliated Hospital of Nanjing Medical University; Geriatric Hospital of Nanjing Medical University, Nanjing, Jiangsu, China.

\*\* Corresponding author.

E-mail addresses: [zhanglouqian@126.com](mailto:zhanglouqian@126.com) (L. Zhang), [ymaent@njmu.edu.cn](mailto:ymaoent@njmu.edu.cn) (Y. Mao).

<sup>1</sup> These authors contributed equally to this work and share the first authorship.

<https://doi.org/10.1016/j.neo.2023.100888>

Received 15 October 2022; Accepted 9 February 2023

1476-5586/© 2023 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

development and progression of NSCLC [6–8]. EGFR mutations have been reported to occur almost exclusively in exons 18 to 21 of the gene, which encodes the EGFR kinase domain [9,10]. About 80–90% of such mutations occur as 19-Del (codons 746 to 754 in exon 19) and L858R (exon 21), termed common EGFR (cEGFR) mutations [11–13]. With extensive analysis of the molecular features of patients harboring cEGFR mutations, accumulating evidence has demonstrated distinguishing clinical characteristics between 19-Del and L858R [4,14]. Moreover, patients with 19-Del or L858R mutations showed differential sensitivity to various chemotherapy or targeted therapies, although some conclusions are still controversial [4,5,14–17].

Despite the high prevalence of cEGFR mutations in NSCLC, approximately 10–20% of cases involve other mutation sites within exons 18 to 21 of EGFR, termed uncommon EGFR (uEGFR) mutations. G719X (exon 18), L861Q (exon 21), S768I (exon 20), and exon 20 insertions (20-Ins) are the most frequent subtypes for uEGFR mutations [18–20]. NSCLC patients with these uEGFR mutations benefit from first-generation (G) EGFR-TKIs such as erlotinib and gefitinib [18–23]. Increasing evidence demonstrated improved outcomes for patients with uEGFR mutations such as G719X, L861Q, and S768I mutations upon treatment with second-generation EGFR-TKIs [23]. On the other hand, EGFR 20-Ins are traditionally believed insensitive to anti-EGFR therapies [24–26]. Overall, data from most retrospective studies of uEGFR mutations is highly heterogeneous. There is no consensus to determine the therapeutic intervention for populations with uEGFR mutations.

This study aims to systemically review the targeted sequencing results of a large cohort of EGFR-mutated NSCLC patients to investigate the molecular landscape of patients with different EGFR mutation subtypes associated with clinical response to first-generation EGFR-TKIs. The genetic-clinical association would provide valuable information to support treatment selection in heterogeneous subgroups of NSCLC patients with rare EGFR mutations.

## Materials and methods

### Patient enrollment and sample collection

Tumor specimens were collected from EGFR-mutated NSCLC patients as a routine diagnosis at all participating hospitals between June 2015 and May 2020. This study was approved by the institutional research ethics committee of the Nanjing Drum Tower Hospital. Written consent was acquired from each patient before sample collection. Qualified samples were subjected to targeted next-generation sequencing by a CLIA-certified and CAP-accredited clinical testing laboratory (Nanjing Geneseeq Technology Inc., Nanjing, China) using pan-cancer gene panels (GENESEEQPRIME™ and RADIOTRON™, Geneseeq Technology Inc.). Formalin-fixed paraffin-embedded (FFPE) tissue samples were confirmed by pathologists from the centralized clinical testing center before genetic testing. Clinical characteristics and treatment history were extracted from medical records. The mutational and corresponding clinical data of the validation cohort comprising 156 EGFR-mutated lung adenocarcinoma patients were downloaded from the cBioPortal for Cancer Genomics (<https://www.cbioportal.org/>) [27,28].

### Targeted next-generation sequencing

DNA extraction, library construction, and targeted capture enrichment were carried out following standard protocols as previously described with modifications [29,30]. FFPE samples were de-paraffinized first with xylene before genomic DNA extraction using QIAamp DNA FFPE Tissue Kit (Qiagen Cat. No. 56404) according to the manufacturer's instructions. Genomic DNA extracted from tumor samples was qualified using Nanodrop2000 (Thermo Fisher Scientific, Waltham, MA), then quantified using the dsDNA HS assay kit on a Qubit 3.0 fluorometer (Life Technology, US) according to the manufacturer's recommendations. Targeted NGS libraries were prepared using the KAPA Hyper

Prep kit (KAPA Biosystems) with an optimized manufacturer's protocol for different sample types. Targeted capture enrichment was performed as previously described [31]. According to the manufacturer's instructions, the target-enriched library was then sequenced on HiSeq4000 and HiSeq4000 NGS platforms (Illumina).

### Mutation calling

Sequencing data was first demultiplexed and subjected to FASTQ file quality control using Trimmomatic [32]. Only data without extra nucleotide bases and passed quality control (QC above 15) were subjected to the following analyses. Raw reads were mapped to the reference Human Genome (hg19) using Burrows-Wheeler Aligner (BWA-mem, v0.7.12; <https://github.com/lh3/bwa/tree/master/bwakit>). Genome Analysis Toolkit (GETK 3.4.0; <https://software.broadinstitute.org/gatk/>) was employed to perform local realignment around the insertions/deletions (INDELS) and base quality score recalibration. Picard was used to remove PCR duplicates. VarScan2 was applied to detect single-nucleotide variations (SNVs) and INDELS. SNVs were filtered out if the mutant allele frequency (MAF) was less than 1% for tumor tissue and 0.3% for plasma samples. Eight out of ten oncogenic signaling pathway alterations were compared among patients with EGFR mutations [33]. The Nrf2 pathway was excluded from the pathway-level analyses of gene alterations co-mutated in EGFR-mutated NSCLC patients, given that KEAP1, NFE2L2, and CUL3 were not included in the 416-cancer-related gene panel. The RTK/RAS pathway was also excluded, given that all patients in the study cohort harbored EGFR mutations.

### Identifying mutation signatures

Information on mutational signatures prevalent in EGFR-mutated NSCLC patients was downloaded from version 3.0 of the Catalog of Somatic Mutations in Cancer (COSMIC) database (<https://cancer.sanger.ac.uk/cosmic/signatures/>) [34,35]. The mutation signature was determined by matching observed mutations to the most frequent nucleotide base changes and the trinucleotide context in each signature. If multiple signatures were compatible with the mutation in question, the most active signature in a specific biological context referring to the scientific literature was used.

### Statistical methods

Plots in this study were generated using the R Project for Statistical Computing (version 3.4.0). Statistical analyses were performed using R. Fisher's exact tests were used to test the categorical variables between groups. Kaplan-Meier curves were used to analyze progression-free survival (PFS) and overall survival (OS) of various patient groups, and the statistical difference was analyzed using the log-rank test. A two-sided P value of less than 0.05 was considered significant for all tests unless indicated otherwise (\* $P < 0.05$ , 0.01 < \*\* $P < 0.05$ , \*\*\* $P < 0.001$ ).

## Results

### Patient overview

2,280 NSCLC patients with known EGFR mutations who underwent targeted NGS of 416 cancer-related genes were included in the following analyses (Supplementary Figs. 1, 2). Patients were further divided into two subgroups depending on whether the patient harbored common or uncommon EGFR mutations at diagnosis. In particular, the cEGFR-positive cohort comprised 1,022 patients with L858R (44.8%) and 973 patients with 19-Del (42.7%). There was no significant difference in the clinical characteristics between these two cohorts (Supplementary Table S1). On the other hand, four uEGFR mutation subtypes were analyzed, including 20-Ins, G719X, L861Q, and S768I. Notably, since many patients with L861Q or S768I harbored co-existing G719X, patients with

**Table 1**  
Clinical characteristics among patients with different EGFR mutation subtypes (N = 2280).

Characteristics	Common N (%)		Uncommon N (%)						P value
	L858R	19-Del	20-Ins	G719X	L861Q	S768I	L861Q+G719X	S768I+G719X	
Gender									
Male	412 (40.3)	417 (42.9)	58 (41.7)	34 (57.6)	15 (33.3)	6 (75.0)	4 (66.7)	16 (57.1)	0.018
Female	610 (59.7)	556 (57.1)	81 (58.3)	25 (42.4)	30 (66.7)	2 (25.0)	2 (33.3)	12 (42.9)	
Age									
≥ 60 years	602 (58.9)	453 (46.6)	69 (49.6)	29 (49.2)	30 (66.7)	3 (37.5)	1 (16.7)	17 (60.7)	<0.001
< 60 years	412 (40.3)	510 (52.4)	70 (50.4)	30 (50.8)	15 (33.3)	4 (50.0)	5 (83.3)	10 (35.7)	
NA	8 (0.8)	10 (1.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (12.5)	0 (0.0)	1 (3.6)	
Histology									
ADC	868 (84.9)	806 (82.8)	120 (86.3)	52 (88.1)	37 (82.2)	5 (62.5)	3 (50.0)	23 (82.1)	0.022
SCC	11 (1.1)	23 (2.4)	5 (3.6)	1 (1.7)	1 (2.2)	1 (12.5)	1 (16.7)	1 (3.6)	
ASC	14 (1.4)	14 (1.4)	2 (1.4)	1 (1.7)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	
NOS	129 (12.6)	130 (13.4)	12 (8.6)	5 (8.5)	7 (15.6)	2 (25.0)	2 (33.3)	4 (14.3)	
Stage at diagnosis									
I	79 (7.7)	64 (6.6)	11 (7.9)	4 (6.8)	2 (4.4)	1 (12.5)	0 (0.0)	0 (0.0)	0.071
II	29 (2.8)	26 (2.7)	4 (2.9)	2 (3.4)	3 (6.7)	0 (0.0)	0 (0.0)	1 (3.6)	
III	41 (4.0)	49 (5.0)	12 (8.6)	6 (10.2)	2 (4.4)	0 (0.0)	0 (0.0)	1 (3.6)	
IV	382 (37.4)	414 (42.6)	57 (41.0)	30 (50.9)	10 (22.2)	3 (37.5)	3 (50.0)	10 (35.7)	
NA	491 (48.0)	420 (43.2)	55 (39.6)	17 (28.8)	28 (62.3)	4 (50.0)	3 (50.0)	16 (57.1)	

P values are based on the Fisher's exact test; ADC, adenocarcinoma; SCC, squamous cell carcinoma; ASC, adenosquamous cell carcinoma; NA: not available; NOS: not otherwise specified; Bold represents significant P values.

double EGFR mutations were separated from patients with single mutations (Supplementary Fig. 2). In brief, the uEGFR-positive cohort contained 139 patients with 20-Ins (6.1%), 59 patients with G719X (2.6%), 45 patients with L861Q (2.0%), 6 patients with L861Q and G719X compound mutations (0.3%), 8 patients with S768I (0.4%), and 28 patients with S768I and G719X compound mutations (1.2%). The demographic and clinical characteristics of patients with varying types of EGFR mutations were summarized in Table 1. A cohort combined using two datasets [27,28] comprising 156 EGFR-mutated patients with NSCLC was used as an external validation set to test the robustness of our mutational profiling results. Corresponding clinical characteristics of these 159 NSCLC patients were summarized in Supplementary Table. S2.

#### Somatic alterations associated with different EGFR mutation subtypes

Since there has been accumulating evidence showing that EGFR mutation subtypes confer different responses to EGFR-TKIs, we analyzed the NGS data of 2,250 EGFR-mutated NSCLC patients and characterized the genomic landscape of these patients to explore the genetic-clinical associations (Fig. 1a). The top frequently co-mutated genes in the cEGFR and uEGFR cohorts include TP53 (55.9% vs. 59.6%), PIK3CA (8.2% vs. 10.5%), RB1 (7.8% vs. 9.1%), CTNNB1 (6.8% vs. 3.5%), APC (6% vs. 6.3%), SMAD4 (5.5% vs. 3.5%) and ARID1A (4.9% vs. 6.7%). Gene-level analyses showed that CTNNB1, which encodes  $\beta$ -catenin protein previously implicated in the pathogenesis of several cancers, was significantly more enriched in patients with cEGFR than those with uEGFR ( $P < 0.05$ , Fig. 1b). In great contrast, genes such as MED12, RICTOR, and GATA1 showed a higher mutation frequency in patients with uEGFR mutations ( $P < 0.01$ ). Pathway-level analyses revealed that hippo pathway gene alterations were more frequently observed in patients harboring uEGFR mutations than those with cEGFR mutations (Fig. 1c). In addition, FAT1 showed a strong correlation with PMS2, MED12, and GATA6 in cEGFR-positive patients ( $P < 0.05$ , Fig. 1d). In contrast, no significant correlation was observed in the uEGFR cohort (Fig. 1e).

We next assessed the tumor mutational burden (TMB) in the cEGFR and uEGFR cohorts. TMB associated with uEGFR mutations was significantly higher than with cEGFR mutations (5.07 vs. 3.80,  $P < 0.01$ , Fig. 1f). Consistent with this finding, a higher TMB was observed in uEGFR-positive patients in the validation cohort ( $P = 0.1746$ , Fig. 1g). We reasoned that the insignificant difference could be due to the limited sample size of validation patients harboring uEGFR mutations (Supplementary Table. S2). Overall, our findings indicate that although patients

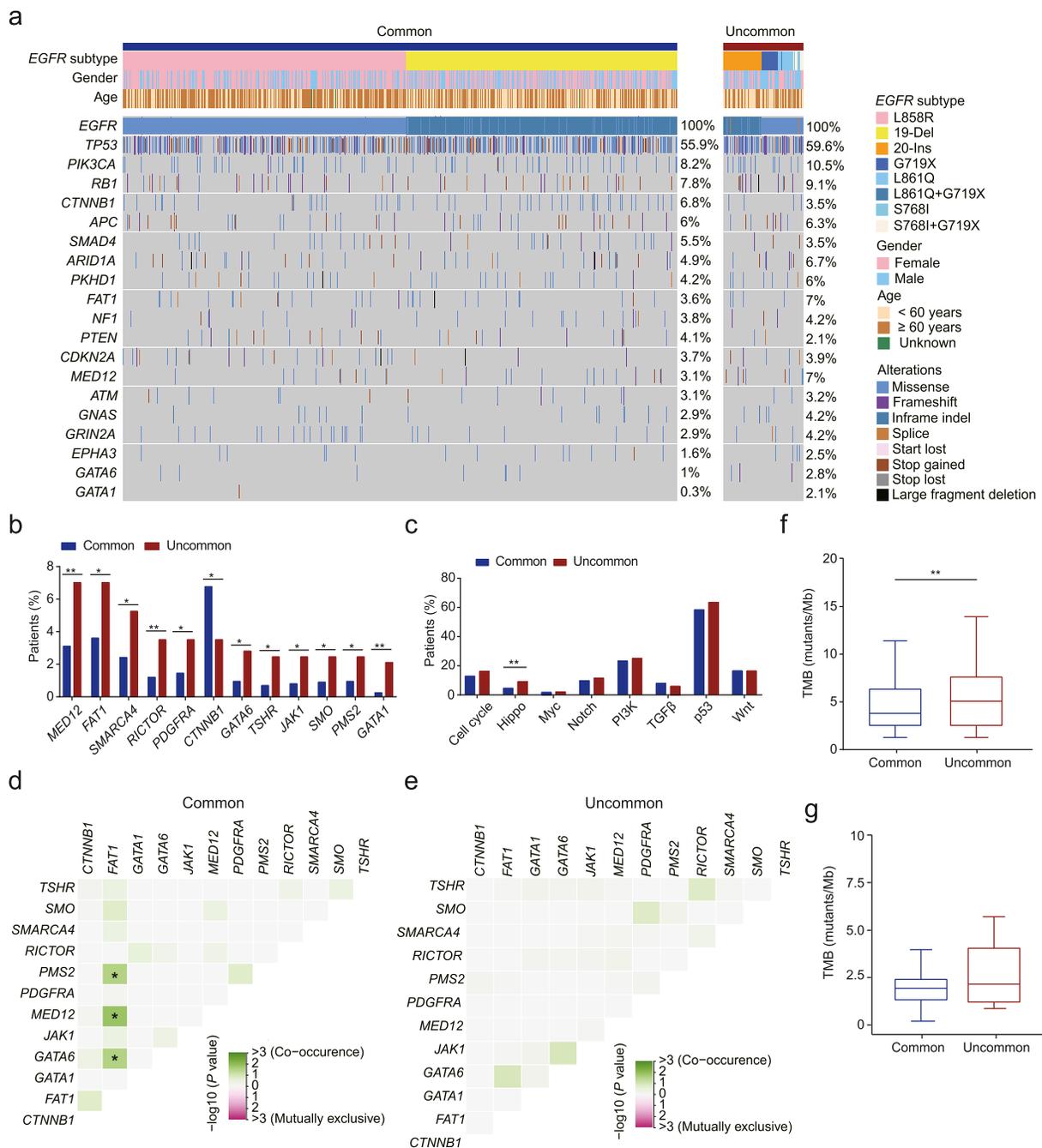
with either cEGFR or uEGFR mutations showed some similarities in their genomic landscapes, for example, high mutational frequency in TP53, PI3KCA, and RB1. Gene- and pathway-level analyses implicated unique molecular features of patients with uEGFR mutations that might be associated with differential responses to targeted therapy.

#### Mutational signatures in patients with common or uncommon EGFR mutations

We then analyzed our data set of somatic mutational signatures associated with common and uncommon EGFR mutations. BRCA1/2-related signature (signature 3) and immunoglobulin-related signature (signature 9) showed a higher prevalence in patients with cEGFR mutations (Fig. 2a). In great contrast, smoking-induced mutations (signature 4) showed enrichment in the uEGFR subgroup, consistent with results obtained by assessing the external validation cohort (Fig. 2b). Unlike findings obtained using the discovery patient cohort, no significant difference in the contribution of mutations in neither the BRCA- nor the immunoglobulin-related signatures was observed, presumably due to the relatively small sample size in cEGFR-positive patients in the validation cohort (1,995 discovery cEGFR vs. 138 validation cEGFR,  $P < 0.001$ , Supplementary Table S2). Overall, these results suggest that NSCLC patients carrying common or uncommon EGFR mutations may have undergone distinct mutational processes, resulting in unique mutational patterns and specific activities in the genome during tumorigenesis and cancer development.

#### Subgroup analysis reveals unique molecular features of EGFR mutation subtypes

Fifty patients in our discovery cohort underwent first-generation EGFR-TKI treatment (Supplementary Table S3). The progression-free survival (PFS) of patients with uEGFR mutations was significantly shorter than those harboring cEGFR mutations (median: 9.03 vs. 4.35 months,  $P = 0.039$ , Fig. 3a), suggesting that NSCLC patients with uEGFR mutations are associated with a poorer prognosis than those with cEGFR mutations on first-generation EGFR-TKI treatment. Consistent with this notion, the overall survival (OS) greatly benefited from first-generation EGFR-TKIs when patients harbored cEGFR mutations rather than uEGFR mutations (median: 85.3 vs. 19.3 months,  $P = 0.01$ , Fig. 3b). Further analyses indicated that none of the clinical characteristics, specifically the patients' gender, age or line of treatment, contributed to the effi-

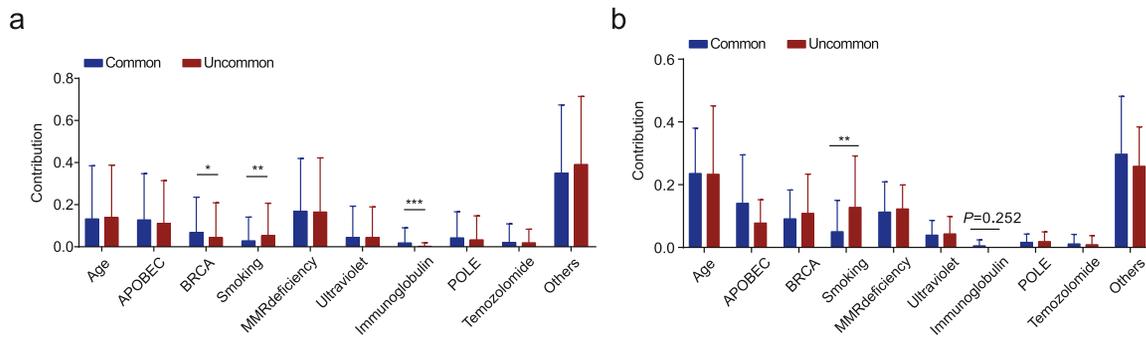


**Fig. 1.** Somatic alterations associated with common and uncommon EGFR mutations. **a.** The genomic landscape of patients with common or uncommon EGFR mutations. Individual gene mutations in baseline tumor samples of EGFR-mutated NSCLC patients ( $N = 2280$ ) were assessed by targeted NGS. Each column represents one patient. Clinical characteristics of patients are shown at the top. The frequency of each gene alteration is listed on the right. **b.** The bar plot shows the different distribution of somatic mutations in the cEGFR and uEGFR patient cohorts. **c.** The correlation between signaling pathways in which the concurrent mutations occur and EGFR subtypes. **d-e.** The heatmap demonstrates how frequently two somatic mutations occur in patients with cEGFR (**d**) or uEGFR mutations (**e**). **f-g.** Higher TMB was more likely associated with patients harboring uEGFR mutations than those carrying cEGFR mutations assessed by using the discovery study cohort (**f**) or the external validation cohort (**g**).

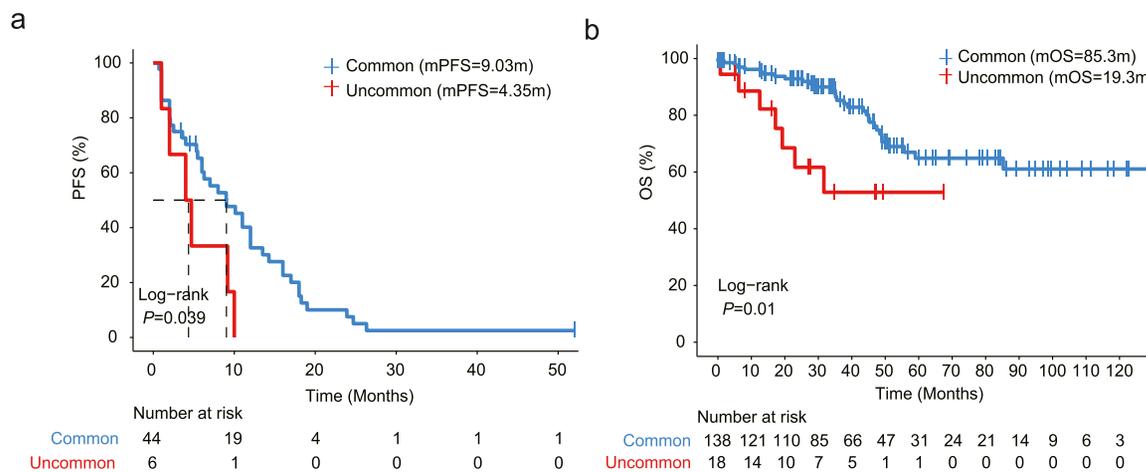
cacy of EGFR-TKIs (Table 2, Supplementary Fig. 3a). However, the PFS of patients with PI3K (median: 5.3 vs. 10.0 months,  $P = 0.047$ ) or TGFβ pathway gene alterations (median: 7.0 vs. 24.7 months,  $P = 0.028$ ) was significantly shorter than their wild-type counterparts (Supplementary Fig. 3b, 3c), implicating that there are other non-negligible factors associated with the differential responses to first-generation EGFR-TKIs in EGFR-mutated NSCLC patients.

*Analysis of molecular features of patients with mutation of 19-Del or L858R*

Since previous studies have demonstrated a prolonged survival of NSCLC patients with EGFR 19-Del than those with L858R [4,15,16], we evaluated the molecular features of patients with either 19-Del or L858R to explain the different clinical responses to EGFR-TKIs. Interestingly, mutations in FLT4, ARID2, MTOR, and DPYD showed a higher preva-



**Fig. 2.** The smoking-related mutational signature was enriched in uEGFR patients. a. Stacked bar plot illustrates the contribution of mutational signatures in cEGFR and uEGFR cohorts using the discovery cohort ( $N = 2,280$ ). b. Stacked bar plot demonstrates the contribution of mutational signatures in cEGFR and uEGFR cohorts using the external validation cohort ( $N = 156$ ).



**Fig. 3.** uEGFR mutations confer a poor prognosis on first-generation EGFR-TKI treatment. a. Kaplan-Meier estimates of progression-free survival (PFS) in NSCLC patients with cEGFR or uEGFR mutations who had received first-generation EGFR-TKI treatment ( $N = 50$ ). b. Kaplan-Meier estimates of overall survival (OS) using the external validation cohort patients ( $N = 156$ ).

**Table 2**  
Univariate analysis of factors associated with first-generation EGFR-TKI clinical responses.

Factor	N (%)	HR for PFS	P value
Gender (female)	27 (54.0)	1.08 (0.60~1.93)	0.805
Age ( $\geq 60$ )	21 (42.0)	1.25 (0.69~2.26)	0.464
EGFR mutation subtype (uEGFR)	6 (12.0%)	2.51 (1.01~6.21)	0.039
Signaling pathways			
Cell cycle	9 (18.0)	0.58 (0.28~1.24)	0.157
Hippo	2 (4.0)	2.66 (0.61~11.52)	0.175
Myc	3 (6.0)	0.74 (0.22~2.51)	0.627
Notch	4 (8.0)	0.97 (0.30~3.16)	0.961
PI3K	11 (22.0)	2.21 (1.00~4.93)	0.047
TGF $\beta$	3 (6.0)	0.22 (0.05~0.94)	0.028
p53	38 (76.0)	1.50 (0.75~3.00)	0.244
Wnt	9 (18.0)	1.32 (0.60~2.88)	0.484

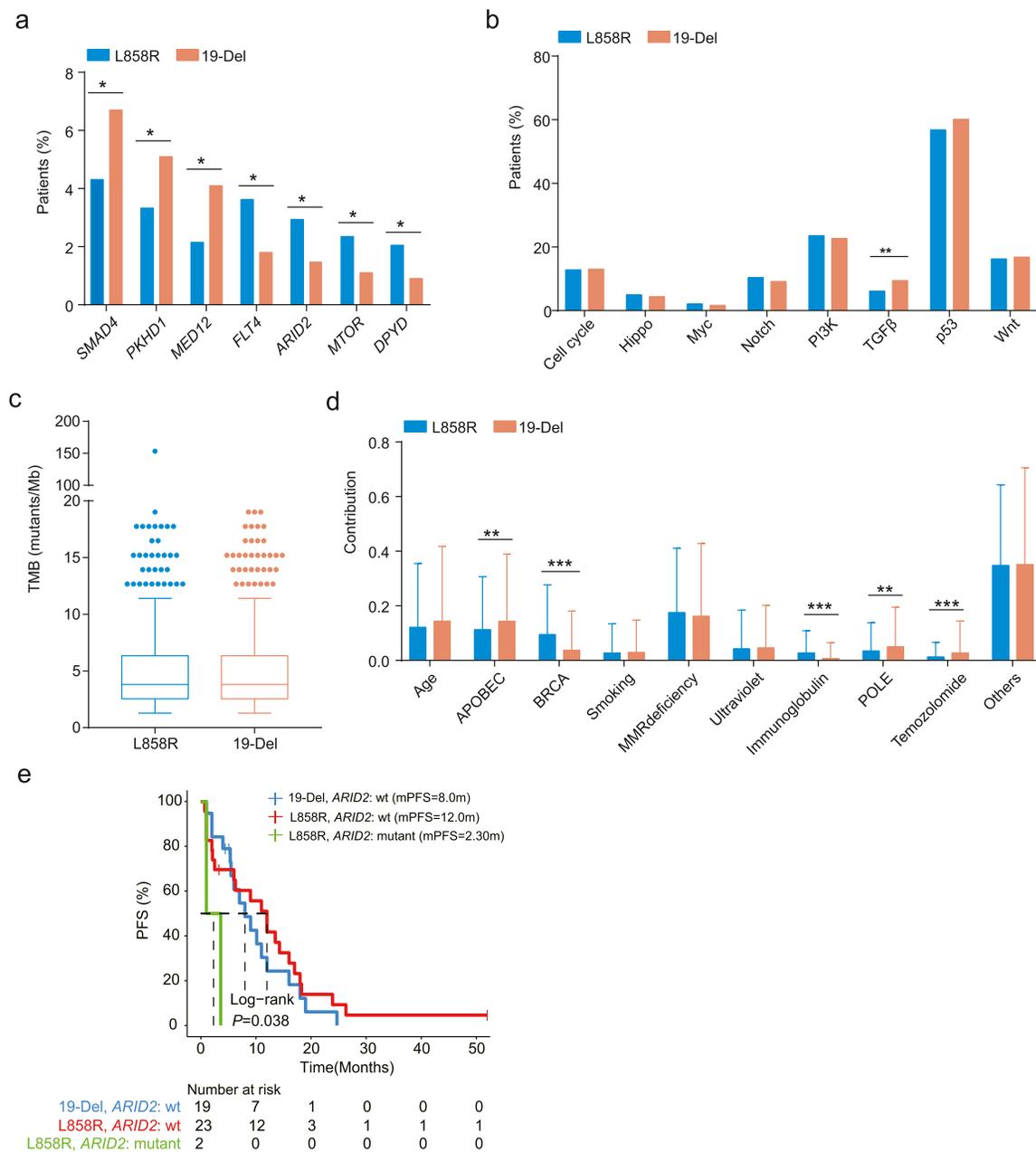
Bold represents statistically significant P values based on the log-rank test.

lence in patients with L858R, whereas mutations in SMAD4, PKHD1, and MED12 were more commonly identified in patients with 19-Del (Fig. 4a). Pathway-level analysis results demonstrated that gene alterations related to the TGF $\beta$  signaling pathway were markedly enriched in patients with 19-Del rather than L858R (Fig. 4b). Although no significant difference in TMB was observed in the 19-Del and L858R cohorts, somatic mutations belonging to the BRCA- and immunoglobulin-related signatures showed a higher prevalence in patients with L858R than those with 19-Del (Fig. 4c, d). Of the four significantly enriched genes in patients with L858R, only ARID2 was associated with patients'

survival after first-generation EGFR-TKI treatment. Specifically, the PFS of patients with L858R-positive and ARID2-positive tumors was significantly shorter than patients with cEGFR-positive and ARID2-negative tumors (Fig. 4e, Supplementary Fig. 3d). Although TGF $\beta$  gene alterations showed a higher prevalence in patients with 19-Del, no significant difference was observed in the PFS of 19-Del-positive patients with or without mutated TGF $\beta$  pathway (median: 16.4 vs. 7.0 months,  $P = 0.17$ , Supplementary Fig. 3e). In addition, the PFS of uEGFR-positive patients carrying TGF $\beta$  gene alterations showed no significant difference compared to uEGFR-positive patients with unaltered TGF $\beta$  pathway ( $P = 0.13$ , Supplementary Fig. 3f).

*Analysis of gene-level and pathway-level changes among EGFR mutation subtypes*

Given that uEGFR mutations in NSCLC are heterogeneous, EGFR-TKIs can have different efficacy in this specific population. Therefore, further assessment is required to decipher the genetic-clinical associations among these patients. Similarly, we assessed gene-level and pathway-level changes among EGFR mutation subtypes, specifically 20-Ins, G719X, L861Q and S768I. Notably, RNF43 mutation was more frequently observed in the G719X cohort, while DICER1, CSF1R, and GATA6 mutations were more likely associated with L861Q (Fig. 5a). Patients with S768I, on the other hand, tended to carry TSHR and EPHA3 mutations compared to other EGFR subtypes. We also showed that mutational signatures were significantly different among EGFR mutation subtypes (Fig. 5b). No significant difference regarding the pathway-level changes or TMB was observed in our analysis (Fig. 5c, d). Lastly, we



**Fig. 4.** Comparing the molecular features of NSCLC patients with L858R or 19-Del mutations. a. Somatic mutations showed differential distribution in patients with L858R and 19-Del mutations. b. Mutations in the TGFβ signaling pathway were enriched in patients with 19-Del. c. No significant difference was observed in TMB between the two cohorts. d. Mutational signatures attributed by L858R and 19-Del mutations. e. Kaplan-Meier estimates of PFS in L858R-positive or 19-Del-positive NSCLC patients with or without ARID2 mutations.

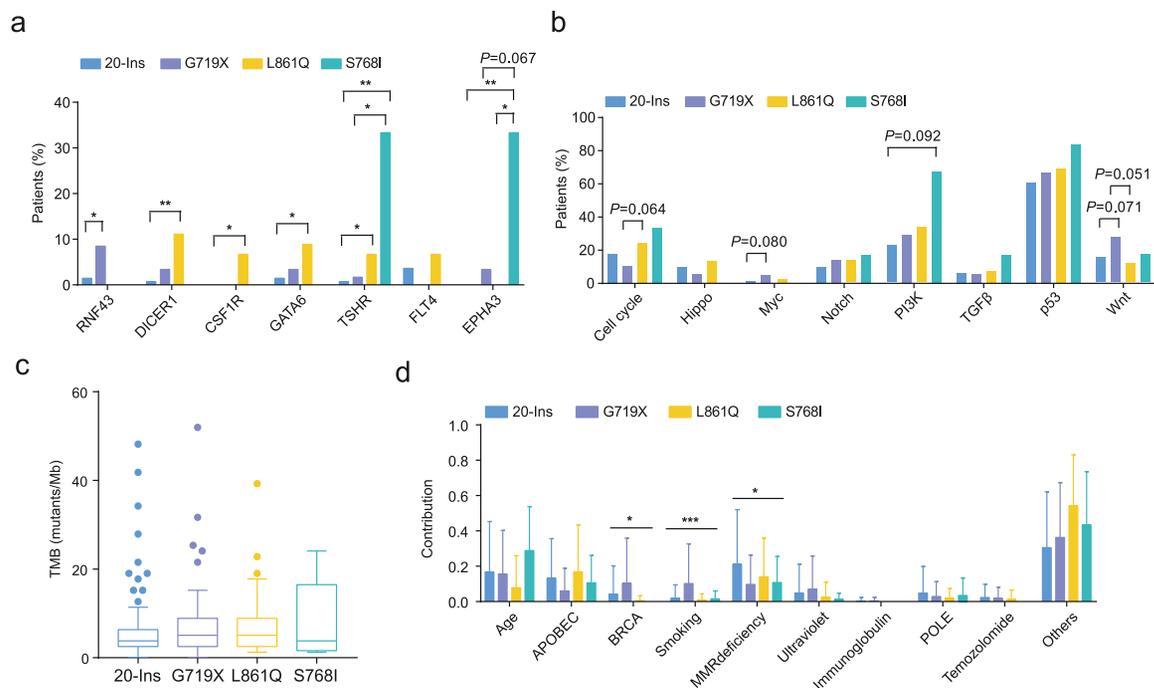
compared the top frequently mutated genes in patients with single and compound EGFR mutations, and no significantly differentially mutated genes were identified (Supplementary Fig. 4). Indeed, previous studies suggest that G719X is frequently observed in patients with L861Q or S768I as a co-existing mutation [36], though the mechanism behind this co-mutation remains unclear.

## Discussion

In this study, we retrospectively analyzed targeted next-generation sequencing results of baseline tumor specimens collected from 2,280 EGFR-mutated patients with NSCLC. We used an external validation cohort comprising 156 patients from two public-available databases to cross-validate our findings. This study indicates that pa-

tients with common or uncommon EGFR mutations have a unique molecular landscape, which probably underlies their differential responses to targeted therapies such as first-generation EGFR-TKIs.

Previous studies have demonstrated the efficacy of EGFR-TKIs on NSCLC patients harboring EGFR 19-Del and L858R mutations. However, very little was known about why the median progression-free survival time of patients harboring 19-Del was significantly improved compared with patients carrying L858R mutation following first-line EGFR-TKI treatment. The current study found that the presence of somatic mutations in ARID2, a gene that encodes a protein that functions in the chromatin remodelling complex to promote gene transcription, might confer a significantly shorter PFS in patients harboring L858R compared to patients with 19-Del (Fig. 4a, e, Supplementary Fig. 3d). We also demonstrated that the enrichment of BRCA1/2- and immunoglobulin-related



**Fig. 5.** Subgroup analysis of the molecular features of NSCLC patients with uEGFR mutations associated with different clinical responses. a. Somatic mutations showed differential distribution among patients with different EGFR mutation subtypes. b. No significant difference was observed in the pathway-level analysis of somatic mutations identified among EGFR mutation subtypes. c. TMB associated with each type of uEGFR mutation showed no significant difference. d. Mutational signatures attributed by uEGFR mutations.

mutational signatures in the L858R cohort might explain the differential sensitivity to EGFR-TKIs (Fig. 4d).

Approximately 10–20% of EGFR mutation sites occur within or, even more rarely, outside the kinase domain of the receptor (i.e., uEGFR mutations). Prior data derived from retrospective series of uEGFR mutations are highly heterogeneous in which the efficacy data are not distinguished between single and compound mutations. An initial objective of this study was to provide explanations for different sensitivity to EGFR-TKIs among patients with various EGFR mutation subtypes by characterizing somatic alterations, TMB and mutational signatures of these patients. This study confirms that uEGFR mutations are associated with a poorer prognosis after first-generation EGFR-TKI treatment (Fig. 3). One possible explanation for this was that patients harboring uEGFR mutations tend to have more Hippo pathway-related gene alterations and higher TMB than those with cEGFR mutations (Fig. 1c, f). Surprisingly, S768I was associated with somatic mutations in TSHR and EPHA3, whereas L861Q was more frequently observed in patients with concurrent DICER1, CSF1R, and GATA6 mutations (Fig. 5a). Unfortunately, this study was unable to identify differential genetic alterations associated with single or compound mutations. One possible explanation for this is that S768I or L861Q shows a lower penetrance in the genome, thereby being more likely to form compound mutations with G719X than other uEGFR mutations.

Despite the above-mentioned promising results, some limitations were associated with the study. For example, due to the limited number of patients with available clinical data, we could not assess the efficacy of EGFR-TKIs in patients harboring EGFR 20-Ins. Previous studies have demonstrated that EGFR 20-Ins-positive patients responded poorly to targeted EGFR inhibitors, including the third-generation EGFR-TKI, osimertinib [37–39]. Future research should be undertaken to investigate why patients with this mutation adopt poor clinical responses and how to improve the survival benefit of these patients.

In conclusion, we systematically elucidated molecular profiles of tumors collected from NSCLC patients harboring different EGFR mutation

subtypes, which showed specific genomic characteristics that might explain differential treatment responses to first-generation EGFR-TKIs.

#### Ethics approval and consent to participate

The study was approved by the Ethical Committee of the Nanjing Drum Tower Hospital (2020-158). Written consent was obtained from each patient.

#### Consent for publication

Written informed consent form was obtained from each patient.

#### Author's contributions

Conception and design: Yongkang Bai, Xiang Liu, Yuan Mao, Louqian Zhang

Collection and assembly of data: Shi Xiong, Pengfei Zhang, Zichen Jiao, Gefei Zhao, Chu Zhou, Junli Zhang, Jiaohui Pang

Data analysis: Yongkang Bai, Xiang Liu, Limin Zheng, Shi Xiong, Pengfei Zhang, Zichen Jiao, Gefei Zhao, Chu Zhou, Song Wang, Junli Zhang, Jiaohui Pang, Yang Xu

Manuscript writing: Yongkang Bai, Xiang Liu, Song Wang, Yang Xu

Study supervision: Yuan Mao, Louqian Zhang

Final approval of manuscript: all authors

#### Funding

This research is supported by the grants from the National Natural Science Foundation (no. 82073389) and Jiangsu Province “333 Project” Research Projects (2022)3-1-327).

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Availability of data and materials

The datasets generated and/or analysed during this current study are available from the corresponding author on reasonable request.

## Acknowledgments

We would like to thank the patients and their families. We are grateful for generous help from investigators and researchers involved in the current study.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neo.2023.100888.

## References

- [1] R.L. Siegel, K.D. Miller, H.E. Fuchs, A. Jemal, Cancer statistics, *CA Cancer J. Clin.* 72 (2022) 7–33.
- [2] D.S. Ettinger, W. Akerley, G. Bepler, M.G. Blum, A. Chang, R.T. Cheney, L.R. Chirieac, T.A. D'Amico, T.L. Demmy, A.K. Ganti, et al., Non-small cell lung cancer, *J. Natl. Compr. Cancer Netw.* 8 (2010) 740–801.
- [3] J.R. Molina, P. Yang, S.D. Cassivi, S.E. Schild, A.A. Adjei, Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship, *Mayo Clin. Proc.* 83 (2008) 584–594.
- [4] C.K. Lee, Y.L. Wu, P.N. Ding, S.J. Lord, A. Inoue, C. Zhou, T. Mitsudomi, R. Rosell, N. Pavlakis, M. Links, et al., Impact of Specific Epidermal Growth Factor Receptor (EGFR) mutations and clinical characteristics on outcomes after treatment with EGFR tyrosine kinase inhibitors versus chemotherapy in egfr-mutant lung cancer: a meta-analysis, *J. Clin. Oncol.* 33 (2015) 1958–1965.
- [5] M. Maemondo, A. I. K. Kobayashi, S. Sugawara, S. Oizumi, H. Isoe, A. Gemma, M. Harada, H. Yoshizawa, I. Kinoshita, Y. Fujita, S. Okinaga, H. Hirano, K. Yoshimori, T. Harada, T. Ogura, M. Ando, H. Miyazawa, T. Tanaka, Y. Saijo, K. Hagiwara, S. Morita, T. Nukiwa, for the North-East Japan Study Group\*, Gefitinib or chemotherapy for non-small-cell lung cancer with mutated EGFR, *N. Engl. J. Med.* (2010).
- [6] W.A. Cooper, D.C. Lam, S.A. O'Toole, J.D. Minna, Molecular biology of lung cancer, *J. Thorac. Dis.* 5 (Suppl 5) (2013) S479–S490.
- [7] F. Ciardiello, G. Tortora, EGFR antagonists in cancer treatment, *N. Engl. J. Med.* 358 (2008) 1160–1174.
- [8] W. Pao, V.A. Miller, K.A. Politi, G.J. Riely, R. Somwar, M.F. Zakowski, M.G. Kris, H. Varmus, Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain, *PLoS Med.* 2 (2005) e73.
- [9] H. Ji, D. Li, L. Chen, T. Shimamura, S. Kobayashi, K. McNamara, U. Mahmood, A. Mitchell, Y. Sun, R. Al-Hashem, et al., The impact of human EGFR kinase domain mutations on lung tumorigenesis and *in vivo* sensitivity to EGFR-targeted therapies, *Cancer Cell* 9 (2006) 485–495.
- [10] R. Sordella, D.W. Bell, D.A. Haber, J. Settleman, Gefitinib-sensitizing EGFR mutations in lung cancer activate anti-apoptotic pathways, *Science* 305 (2004) 1163–1167.
- [11] T. Kosaka, Y. Yatabe, H. Endoh, H. Kuwano, T. Takahashi, T. Mitsudomi, Mutations of the epidermal growth factor receptor gene in lung cancer: biological and clinical implications, *Cancer Res.* 64 (2004) 8919–8923.
- [12] W. Pao, V. Miller, M. Zakowski, J. Doherty, K. Politi, I. Sarkaria, B. Singh, R. Heelan, V. Rusch, L. Fulton, et al., EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated with sensitivity of tumors to gefitinib and erlotinib, *Proc. Natl. Acad. Sci. U. S. A.* 101 (2004) 13306–13311.
- [13] J.G. Paez, P.A. Janne, J.C. Lee, S. Tracy, H. Greulich, S. Gabriel, P. Herman, F.J. Kaye, N. Lindeman, T.J. Boggon, et al., EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy, *Science* 304 (2004) 1497–1500.
- [14] Y. Zhang, D. He, W. Fang, S. Kang, G. Chen, S. Hong, J. Sheng, J. Zhan, N. Chen, Z. Hu, et al., The difference of clinical characteristics between patients with Exon 19 deletion and those with L858R mutation in non-small cell lung cancer, *Medicine* 94 (2015) e1949 (Baltimore).
- [15] W. Hong, Q. Wu, J. Zhang, Y. Zhou, Prognostic value of EGFR 19-del and 21-L858R mutations in patients with non-small cell lung cancer, *Oncol. Lett.* 18 (2019) 3887–3895.
- [16] D.W. Jackman, B.Y. Yeap, L.V. Sequist, N. Lindeman, A.J. Holmes, V.A. Joshi, D.W. Bell, M.S. Huberman, B. Halmos, M.S. Rabin, et al., Exon 19 deletion mutations of epidermal growth factor receptor are associated with prolonged survival in non-small cell lung cancer patients treated with gefitinib or erlotinib, *Clin. Cancer Res.* 12 (2006) 3908–3914.
- [17] G.J. Riely, W. Pao, D. Pham, A.R. Li, N. Rizvi, E.S. Venkatraman, M.F. Zakowski, M.G. Kris, M. Ladanyi, V.A. Miller, Clinical course of patients with non-small cell lung cancer and epidermal growth factor receptor exon 19 and exon 21 mutations treated with gefitinib or erlotinib, *Clin. Cancer Res.* 12 (2006) 839–844.
- [18] J. Shi, H. Yang, T. Jiang, X. Li, C. Zhao, L. Zhang, S. Zhao, X. Liu, Y. Jia, Y. Wang, et al., Uncommon EGFR mutations in a cohort of Chinese NSCLC patients and outcomes of first-line EGFR-TKIs and platinum-based chemotherapy, *Chin. J. Cancer Res.* 29 (2017) 543–552.
- [19] Y. Zhang, Z. Wang, X. Hao, X. Hu, H. Wang, Y. Wang, J. Ying, Clinical characteristics and response to tyrosine kinase inhibitors of patients with non-small cell lung cancer harboring uncommon epidermal growth factor receptor mutations, *Chin. J. Cancer Res.* 29 (2017) 18–24.
- [20] J. Xu, B. Jin, T. Chu, X. Dong, H. Yang, Y. Zhang, D. Wu, Y. Lou, X. Zhang, H. Wang, et al., EGFR tyrosine kinase inhibitor (TKI) in patients with advanced non-small cell lung cancer (NSCLC) harboring uncommon EGFR mutations: a real-world study in China, *Lung Cancer* 96 (2016) 87–92.
- [21] S. Watanabe, Y. Minegishi, H. Yoshizawa, M. Maemondo, A. Inoue, S. Sugawara, H. Isoe, M. Harada, Y. Ishii, A. Gemma, et al., Effectiveness of gefitinib against non-small-cell lung cancer with the uncommon EGFR mutations G719X and L861Q, *J. Thorac. Oncol.* 9 (2014) 189–194.
- [22] P. Krawczyk, D.M. Kowalski, R. Ramlaou, E. Kalinka-Warzocho, K. Winiarczyk, K. Stencel, T. Powrozek, K. Reszka, K. Wojas-Krawczyk, M. Bryl, et al., Comparison of the effectiveness of erlotinib, gefitinib, and afatinib for treatment of non-small cell lung cancer in patients with common and rare EGFR gene mutations, *Oncol. Lett.* 13 (2017) 4433–4444.
- [23] Y.L. Wu, V. Hirsh, L.V. Sequist, C.P. Hu, J. Feng, S. Lu, Y. Huang, M. Schuler, T. Mok, N. Yamamoto, et al., Does EGFR mutation type influence patient-reported outcomes in patients with advanced EGFR mutation-positive non-small-cell lung cancer? Analysis of two large, Phase III studies comparing afatinib with chemotherapy (LUX-Lung 3 and LUX-Lung 6), *Patient* 11 (2018) 131–141.
- [24] G.R. Oxnard, P.C. Lo, M. Nishino, S.E. Dahlberg, N.I. Lindeman, M. Butaney, D.M. Jackman, B.E. Johnson, P.A. Janne, Natural history and molecular characteristics of lung cancers harboring EGFR exon 20 insertions, *J. Thorac. Oncol.* 8 (2013) 179–184.
- [25] H. Greulich, T.H. Chen, W. Feng, P.A. Janne, J.V. Alvarez, M. Zappaterra, S.E. Bulmer, D.A. Frank, W.C. Hahn, W.R. Sellers, et al., Oncogenic transformation by inhibitor-sensitive and -resistant EGFR mutants, *PLoS Med.* 2 (2005) e313.
- [26] J.Y. Wu, S.G. Wu, C.H. Yang, C.H. Gow, Y.L. Chang, C.J. Yu, J.Y. Shih, P.C. Yang, Lung cancer with epidermal growth factor receptor exon 20 mutations is associated with poor gefitinib treatment response, *Clin. Cancer Res.* 14 (2008) 4877–4882.
- [27] J. Chen, H. Yang, A.S.M. Teo, L.B. Amer, F.G. Sherbaf, C.Q. Tan, J.J.S. Alvarez, B. Lu, J.Q. Lim, A. Takano, et al., Genomic landscape of lung adenocarcinoma in East Asians, *Nat. Genet.* 52 (2020) 177–186.
- [28] Cancer Genome Atlas Research N, Comprehensive molecular profiling of lung adenocarcinoma, *Nature* 511 (2014) 543–550.
- [29] Z. Yang, N. Yang, Q. Ou, Y. Xiang, T. Jiang, X. Wu, H. Bao, X. Tong, X. Wang, Y.W. Shao, et al., Investigating novel resistance mechanisms to third-generation EGFR tyrosine kinase inhibitor osimertinib in non-small cell lung cancer patients, *Clin. Cancer Res.* 24 (2018) 3097–3107.
- [30] Y. Shu, X. Wu, X. Tong, X. Wang, Z. Chang, Y. Mao, X. Chen, J. Sun, Z. Wang, Z. Hong, et al., Circulating tumor DNA mutation profiling by targeted next generation sequencing provides guidance for personalized treatments in multiple cancer types, *Sci. Rep.* 7 (2017) 583.
- [31] W. Fang, Y. Ma, J.C. Yin, S. Hong, H. Zhou, A. Wang, F. Wang, H. Bao, X. Wu, Y. Yang, et al., Comprehensive Genomic profiling identifies novel genetic predictors of response to Anti-PD-(L)1 therapies in non-small cell lung cancer, *Clin. Cancer Res.* 25 (2019) 5015–5026.
- [32] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, *Bioinformatics* 30 (2014) 2114–2120.
- [33] F. Sanchez-Vega, M. Mina, J. Armenia, W.K. Chatila, A. Luna, K.C. La, S. Dimitriadou, D.L. Liu, H.S. Kantheti, S. Saghafein, et al., Oncogenic signaling pathways in the cancer genome atlas, *Cell* 173 (2018) 321–337 e310.
- [34] J.G. Tate, S. Bamford, H.C. Jubb, Z. Sondka, D.M. Beare, N. Bindal, H. Boutselakis, C.G. Cole, C. Creatore, E. Dawson, et al., COSMIC: the catalogue of somatic mutations in cancer, *Nucleic Acids Res.* 47 (2019) D941–D947.
- [35] C. Lefebvre, T. Bachelot, T. Filleron, M. Pedrero, M. Campone, J.C. Soria, C. Massard, C. Levy, M. Arnedos, M. Lacroix-Triki, et al., Mutational profile of metastatic breast cancers: a retrospective analysis, *PLoS Med.* 13 (2016) e1002201.
- [36] J.H. Cho, S.H. Lim, H.J. An, K.H. Kim, K.U. Park, E.J. Kang, Y.H. Choi, M.S. Ahn, M.H. Lee, J.M. Sun, et al., Osimertinib for patients with non-small-cell lung cancer harboring uncommon EGFR mutations: a multicenter, open-label, Phase II Trial (KCSG-LU15-09), *J. Clin. Oncol.* 38 (2020) 488–495.
- [37] Y. Kobayashi, T. Mitsudomi, Not all epidermal growth factor receptor mutations in lung cancer are created equal: perspectives for individualized treatment strategy, *Cancer Sci.* 107 (2016) 1179–1186.
- [38] B. Van Veggel, A. Van Der Wekken, S. Hashemi, R. Cornelissen, K. Monkhorst, D. Heidemant, T. Radonic, E. Schuurings, E. Smit, J. De Langen, P2.13-42 osimertinib treatment for patients with EGFR exon 20 insertion positive non-small-cell lung cancer, *J. Thorac. Oncol.* 13 (2018) S815.
- [39] S. Vyse, P.H. Huang, Targeting EGFR exon 20 insertion mutations in non-small cell lung cancer, *Signal Transduct. Target. Ther.* 4 (2019) 5.