Original Article

Molecular and clinical disparity of *EGFR*-mutant non-small cell lung cancer (NSCLC) based on histopathological stage and *EGFR* molecular subtypes

Dayoung Yoon[1,2,3], Ji Won Lee[4], Byoung Chul Cho[5], Eun Joo Kang[6], Jung Sun Kim[7], Taekyu Lim[8], Seong Yoon Yi[9], Yu Jung Kim[10], Mi Sun Ahn[11], Young Saing Kim[12], Ji Hyun Park[13], Seungtaek Lim[14], Hyung Soon Park[15], Jang Ho Cho[16], Byunghyun Jang[1,2], Ji Yoon Lee[1,2], Jiwon Kim[1,2], Jisoo Hong[17], Harim Koo[1,2], Seok Chung[3,18], Sang Won Shin[4], Yeul Hong Kim[4], Jason K. Sa[1,2,17], Yoon Ji Choi[4,17]

[1]Department of Biomedical Informatics, Korea University College of Medicine, Seoul, Korea; [2]Department of Biomedical Sciences, Korea University College of Medicine, Seoul, Korea; [3]KU-KIST Graduate School of Converging Science and Technology, Korea University, Seoul, Korea; [4]Division of Hematology-Oncology, Department of Internal Medicine, Korea University Anam Hospital, Korea University College of Medicine, Seoul, Korea; [5]Division of Medical Oncology, Yonsei Cancer Center, Seoul, Korea; [6]Division of Hematology-Oncology, Department of Internal Medicine, Korea University Guro Hospital, Korea University College of Medicine, Seoul, Korea; [7]Division of Hematology-Oncology, Department of Internal Medicine, Korea University Ansan Hospital, Korea University College of Medicine, Ansan, Korea; [8]Division of Hematology-Oncology, Department of Internal Medicine, Veterans Health Service Medical Center, Seoul, Korea; [9]Division of Hematology-Oncology, Department of Internal Medicine, Inje University Ilsan Hospital, Goyang, Korea; [10]Department of Internal Medicine, Bundang Hospital, Seoul National University College of Medicine, Seongnam, Korea; [11]Ajou University Medical Center, Suwon, Korea; [12]Division of Medical Oncology, Department of Internal Medicine, Gil Medical Center, Gachon University College of Medicine, Incheon, Korea; [13]Department of Hemato-Oncology, Konkuk Medical Center, University of Konkuk College of Medicine, Seoul, Korea; [14]Department of Oncology, Wonju Severance Christianity Hospital, Wonju-si, Korea; [15]Division of Medical Oncology, Department of Internal Medicine, St. Vincent's Hospital, College of Medicine, The Catholic University of Korea, Suwon, Korea; [16]Division of Oncology, Department of Internal Medicine, Incheon St. Mary's Hospital, The Catholic University of Korea, Seoul, Korea; [17]oncoMASTER Inc., Seoul, Korea; [18]School of Mechanical Engineering, College of Engineering, Korea University, Seoul, Korea

Contributions: (I) Conception and design: JK Sa, YJ Choi; (II) Administrative support: None; (III) Provision of study materials or patients: JW Lee, BC Cho, EJ Kang, JS Kim, T Lim, SY Yi, YJ Kim, MS Ahn, YS Kim, JH Park, S Lim, HS Park, JH Cho, SW Shin, YH Kim; (IV) Collection and assembly of data: JW Lee, BC Cho, EJ Kang, JS Kim, T Lim, SY Yi, YJ Kim, MS Ahn, YS Kim, JH Park, S Lim, HS Park, JH Cho, SW Shin, YH Kim, B Jang, JY Lee, J Kim, J Hong, H Koo, S Chung; (V) Data analysis and interpretation: D Yoon, JK Sa, YJ Choi; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Jason K. Sa, PhD. Department of Biomedical Informatics, Korea University College of Medicine, 73 Goryeodae-ro, Seongbuk-gu, Seoul 02841, Korea; Department of Biomedical Sciences, Korea University College of Medicine, Seoul, Korea; oncoMASTER Inc., Seoul, Korea. Email: jasonksa@korea.ac.kr; Yoon Ji Choi, MD, PhD. Division of Hematology-Oncology, Department of Internal Medicine, Korea University Anam Hospital, Korea University College of Medicine, 73 Goryeodae-ro, Seongbuk-gu, Seoul 02841, Korea; oncoMASTER Inc., Seoul, Korea. Email: yj_choi@korea.ac.kr.

Background: While epidermal growth factor receptor (EGFR) tyrosine kinase inhibitors (TKIs) are a cornerstone of therapy for advanced *EGFR*-mutant non-small cell lung cancer (NSCLC), resistance remains a major clinical challenge. The genomic landscape of early-stage (ES) *EGFR*-mutant NSCLC and its evolution to advanced-stage (AS) disease is not fully understood. This study aimed to characterize the molecular disparities between ES and AS *EGFR*-mutant NSCLC and to identify genomic alterations associated with EGFR-TKI treatment outcomes.

Methods: We have collected and profiled the complex genomes of 121 ES and 74 AS NSCLCs to determine their molecular and clinical disparities. Furthermore, we analyzed 84 *EGFR*-mutant NSCLC patients who were treated with EGFR-TKIs to identify potential molecular correlates that could predict the treatment response within the clinic. Patients were stratified by progression-free survival (PFS) and overall response rate (ORR), and hazard ratio analyses were performed.

Results: In the study, significant enrichment of mutations in *MTOR*, *ATRX*, *STAG2*, *ABL1*, and *SPEN* was

observed in AS tumors, whereas ES tumors predominantly exhibited mutations activating *JAK2*, *ERBB2*, and *FGFR4*. In the EGFR-TKI cohort, poor responders harbored frequent mutations in *TP53*, *KIT*, and *ALK*, and these were associated with worse clinical outcomes. Conversely, favorable responders showed enrichment of *MTOR*, *ATM*, *EP300*, and *PIK3R1* mutations. *ALK* and *FANCA* were linked to increased hazard, while *EP300* and *PIK3R1* mutations correlated with improved prognosis.

Conclusions: Given the growing importance of biomarker-driven treatment in the field of oncology, our results collectively open up new therapeutic opportunities for ES NSCLC patients.

Keywords: Non-small cell lung cancer (NSCLC); epidermal growth factor receptor mutations (*EGFR* mutations); tyrosine kinase inhibitor-treatments (TKI-treatments); stage; TKI resistance

Submitted Nov 26, 2025. Accepted for publication Jan 22, 2026. Published online Feb 26, 2026.
doi: 10.21037/tlcr-2025-1-1354
View this article at: https://dx.doi.org/10.21037/tlcr-2025-1-1354

Introduction

Comprehensive molecular profiling of the tumor has enabled the exploration of various treatment opportunities for cancer patients with unique molecular abnormalities, demonstrating the clinical feasibility of precision oncology (1). In an effort to acquire reliable genetic information of individual patients, next-generation sequencing (NGS) has revolutionized the field of genomics by producing a massive high-throughput sequencing of the entire genome or targeted regions of interest. Especially within the field of oncology, NGS has demonstrated remarkable success in identifying essential genomic aberrations and biomarkers that can guide and predict patient clinical outcomes to various treatments across a broad range of different tumor types, including breast adenocarcinoma, colorectal adenocarcinoma, and non-small cell lung cancer (NSCLC) (2,3).

NSCLC is the most common malignant form of lung cancer, comprising over 85% of all lung cancer patients (4-7). NSCLC manifests profound levels of genomic aberrations that can directly lead to patient-tailored treatment opportunities; hence, targeted therapies have been the fundamental component of NSCLC treatment. For example, among various well-known proto-oncogenes, epidermal growth factor receptor (*EGFR*) activates a series of intracellular signaling programs that promote cellular proliferation, invasion, and survival. Moreover, genomic alterations in the *EGFR* gene, specifically through mutations, have been identified as one of the most crucial drivers of genomic events, especially in East Asian populations (5,8,9). *EGFR* mutations are often detected at various "hot-spot" domains, including the substitution of leucine to arginine (L858R) in exon 21 and small exon 19 deletions (Del19), both of which are located within the kinase domains (9). The activating mutations within these domains often lead to hypersensitivity to various EGFR tyrosine kinase inhibitors (TKIs), such as erlotinib, gefitinib, afatinib, and osimertinib (10,11). Interestingly,

Highlight box

Key findings

- This study identified dynamic molecular disparity between early-stage (ES) and advanced-stage (AS) epidermal growth factor receptor (*EGFR*)-mutant non-small cell lung cancer (NSCLC). Stage-specific differences were observed in co-occurring driver mutations, pathway alterations, and mutational signatures. In addition, specific genomic alterations were associated with differential clinical responses to EGFR tyrosine kinase inhibitors (TKIs).

What is known and what is new?

- *EGFR* mutations are well-established drivers in NSCLC, particularly in East Asian populations, and EGFR-TKIs have been standard therapy in advanced disease. However, direct genomic comparisons between ES and AS *EGFR*-mutant NSCLC have been limited.
- This study provides a stage-based comparative analysis using next-generation sequencing platform while stratifying tumors by *EGFR* mutation subtypes, revealing profound molecular heterogeneity.

What is the implication, and what should change now?

- The findings highlight the importance of stage-specific genomic context in *EGFR*-mutant NSCLC and suggest that distinct molecular mechanisms may underlie tumor progression and treatment response. These results support the integration of comprehensive genomic profiling into clinical decision-making and future biomarker-driven studies across disease stages.

previous studies have shown that different variations of *EGFR* mutations constituted diverse treatment responses. For example, Del19 prompted diminished levels of phosphotyrosine, whereas the phosphorylation of the L858R mutation sensitized tumor cells more susceptible to erlotinib and gefitinib treatment (12). Furthermore, osimertinib, a third-generation EGFR-TKI, has demonstrated remarkable clinical response in patients with advanced-stage (AS) NSCLC compared to first-generation EGFR-TKIs (11).

NSCLCs can be classified according to various stages based on tumor size, anatomical location, histopathological characteristics, and the extent of tumor spread. For the treatment of NSCLC, if a complete surgical resection doesn't seem viable, patients are often referred to non-invasive clinical management composed of various treatment modalities, including chemotherapy, radiotherapy, immunotherapy, and targeted therapy. Based on such procedure, a majority of the *EGFR*-mutant patients with AS NSCLC often receive EGFR-TKIs as the first-line treatment, whereas early-stage (ES) NSCLC patients are often subjected to surgical resection as the first-line treatment (9,13). However, more than 50% of these patients eventually undergo tumor relapse, and the pathological stage has been identified as the best predictor of outcome (13,14). Due to surgical intervention, the molecular and functional impacts of *EGFR* mutation in ES NSCLCs remain relatively elusive. Previous studies have largely focused on the molecular characteristics of AS NSCLC, and only more recently have a limited number of studies begun to explore ES disease. Consequently, relatively few studies have directly compared the molecular and clinical features of ES and AS NSCLC. Moreover, such comparisons have rarely been performed using the NGS platform while simultaneously stratifying tumors by *EGFR* mutation subtypes. In this study, we address this gap by conducting a stage-based comparative analysis within a single, consistently sequenced cohort of *EGFR*-mutant NSCLC patients.

Therefore, in this study, we collected and characterized the complex genomes of 195 lung cancer patients to explore and unravel the dynamic molecular properties of *EGFR*-mutant NSCLCs at ES and AS. As a result, we discovered that ES and AS NSCLC tumors were characterized by unique genomic aberrations according to various *EGFR* domains. Lastly, *EGFR* mutations at distinct domains manifested diverse clinical outcomes, highlighting the molecular and clinical heterogeneity of NSCLCs. We present this article in accordance with the REMARK reporting checklist (available at https://tlcr.amegroups.com/article/view/10.21037/tlcr-2025-1-1354/rc).

Methods

K-MASTER project and tumor specimen collection

The K-MASTER project is a government-supported precision medicine initiative that focuses on the identification and clinical management of cancer patients. The primary objective of the program is to collect and analyze comprehensive genomic data from 10,000 Korean patients with advanced solid tumors enrolled in the KM-00 master screening protocol. Based on the results of the KM-00 screening, K-MASTER has initiated 20 clinical trials investigating single-agent or combination targeted therapies. The study was conducted in accordance with the Declaration of Helsinki and its subsequent amendments. Following the approval by the Ethics Committee of Korea University Anam Hospital (approval No. 2017AN0401), patients with advanced solid tumors were enrolled in the KM-00 master screening protocol at one of 55 participating institutions. Korea University Anam Hospital served as the lead, and all participating hospitals were informed and agreed to the study. Written informed consent was obtained from all patients prior to sample collection.

Tumor tissue samples were collected from enrolled patients and used for NGS analysis. In addition, liquid biopsy samples were obtained from peripheral blood, and circulating tumor DNA (ctDNA) was extracted for NGS analysis. Tumor tissue and liquid biopsy samples were processed, and sequencing was performed shortly after sample collection.

To support clinical decision-making regarding the actionability of genomic alterations, a Match Master System was developed using the OncoKB knowledge base. Updated evidence from major clinical studies was incorporated into this system.

Sample group according to EGFR subtypes

EGFR mutations were classified into common and uncommon subtypes. Common *EGFR* mutations included Del19 and the L858R point mutation. Uncommon *EGFR* mutations were defined as all other *EGFR* alterations not included in the common mutation category.

For EGFR-TKI response analyses, patients who

received EGFR-TKI treatment were selected from the overall cohort. In addition, to increase the sample size for pharmacogenomic evaluation, additional *EGFR*-mutant NSCLC patients with available treatment and outcome data were included in the treatment response cohort.

Variant calling

Sequencing reads in FASTQ format were aligned to the human reference genome (hg19) using the Burrows-Wheeler Aligner. The resulting BAM files underwent standard preprocessing steps, including sorting, duplicate read removal, local realignment around small insertions and deletions, and base quality score recalibration using SAMtools, Picard, and the Genome Analysis Toolkit (GATK). Somatic variants were identified using MuTect2 to generate high-confidence mutation calls. Known polymorphic sites were filtered using reference databases, including the 1000 Genomes Project, gnomAD, and dbSNP. Variant annotation was performed using the Variant Effect Predictor. Variants with a minimum sequencing depth of ≥20 and a variant allele frequency of ≥2% were retained for downstream analyses.

Mutational signatures

Mutational signature analysis was performed using the deconstructSigs package in R. Six base substitution subtypes (C>T, C>A, C>G, T>C, T>A, and T>G) were used to characterize the mutational profiles of each tumor. The analysis decomposes the observed mutational spectrum into contributions from a predefined set of reference mutational signatures.

Reference signatures were derived from the COSMIC mutational signature database (signature.cosmic). The following signatures were included in the analysis based on prior characterization: age-related (signature 1), defective homologous recombination DNA repair (signature 3), smoking-associated (signature 4), unknown etiology (signatures 8, 12, 17, and 19), defective DNA mismatch repair (signatures 15 and 21), aflatoxin-associated (signature 24), and base excision repair deficiency (signature 30).

Only mutational signatures contributing to at least 7% of the total mutations in a given sample were included in downstream analyses to reduce noise and focus on predominant mutational processes.

Clinical trial enrollment

Genetic alterations, including SNVs, indels, CNAs, or structural rearrangements with clinical actionability, were reported in a clinical report format. Treatment options, including clinical trials in the K-MASTER program, were recommended based on the OncoKB knowledge database (OncoKB API; December 2020) and inclusion criteria for each trial.

ES and AS NSCLC pathway diversity

PathwayMapper (pathwaymapper.org) → pathway diagrams and gene curated and provide tools → major oncogenic signature in The Cancer Genome Atlas (TCGA) journal.

Statistical analyses

Comparisons between groups were performed using the Student's *t*-test, Wilcoxon rank-sum test, or Fisher's exact test, as appropriate. Survival analyses were conducted using the Kaplan-Meier method, with differences assessed by the log-rank test, and hazard ratios were estimated using the Cox proportional hazards regression model. Patients who were alive at the time of last follow-up were censored in survival analyses. All statistical analyses were performed using R software version 4.1.3 (https://www.r-project.org).

Results

Comprehensive molecular landscape of ES and AS EGFR-mutant NSCLC

EGFR mutations represent one of the most significant oncogenic drivers in East Asian patients with lung cancer (8). To investigate stage-specific molecular characteristics of *EGFR*-mutant NSCLC, we retrospectively collected and analyzed tumor specimens from 195 patients diagnosed with NSCLC who harbored somatic *EGFR* mutations to characterize stage-based genomic features. All samples were subjected to NGS, and the clinical and molecular characteristics of this cohort are summarized in Table S1. Among these patients, a subset received EGFR-TKI therapy and was included in the treatment response analysis. To maximize statistical power for pharmacogenomic analyses, additional *EGFR*-mutant NSCLC patients treated with EGFR-TKIs were included in the response cohort,

as detailed in the Methods section. Patients have been categorized into two classifications, and those who have been diagnosed with stage I–IIIA, termed ES, and those with IIIB–IV as AS (9). ES NSCLC patients constituted 62% of our cohort (n=121), while the number of AS NSCLC patients was 74. The major histopathological characteristics of our cohort were predominantly lung adenocarcinoma (LUAD; n=170, 87.2%), followed by poorly differentiated carcinoma (PD; n=25, 12.8%; Figure S1A). We observed a significant difference between ES and AS patients, with poorly differentiated tumors being more frequently observed in the AS group (21.6% *vs.* 7.4%; Figure S1A). Consistent with disease stage, overall survival was significantly shorter in AS patients compared with ES patients, as demonstrated by Kaplan-Meier survival analysis (Figure S1B). As we specifically focused on *EGFR*-mutant NSCLCs, *EGFR* mutation was the most frequently altered gene that was detected across various domains. Among them, mutations in the L858R domain at exon 21 and small Del19 were the most prevalent genomic events (*Figure 1A*) (8,15). The next most frequently mutated gene was *TP53* in both groups, observed at 50% and 43% in the ES and AS groups, respectively. When we determined the frequency of major cancer driver mutations (genes that are frequently identified in cancer, including *TP53*, *PTEN*, and *RB1*) between ES and AS patients, mutations in *ATR*, *ATRX*, *MTOR*, and *ERBB2* demonstrated the most robust differences (*Figure 1B*). Notably, *MTOR* mutation was only detected in the AS tumors, while ES NSCLCs showed higher levels of *ERBB2*, *FGFR4*, and *JAK2* mutations (*Figure 1C*). The mutational signature analysis further delineated the potential etiology of ES and AS tumors. We focused specifically on signature activities that constituted over 7% of the patients in each group, as well as signature 4, which was associated with smoking. Among them, mutational signatures 1 (age), 17 (unknown), 24 (aflatoxin), and 30 (deficiency in base excision pair) exhibited significant differences between the two groups. Although aflatoxin-associated mutational signature (signature 24) has been primarily described in hepatocellular carcinoma, prior epidemiological and experimental evidence suggests that aflatoxin B1 exposure may also contribute to lung carcinogenesis through mutagenic and oncogenic signaling mechanisms (16,17). While ES tumors showed higher levels of base excision repair and aflatoxin-associated signatures, AS tumors were marked by enrichments of signature activities that were associated with spontaneous deamination of 5-methylcytosine and DNA damage by reactive oxygen species (ROS) (*Figure 1D*). As previous studies have reported the synergistic effects in the generation of ROS due to tobacco smoking, we suspected that AS patients may have largely propagated from exposure to smoking, while ES patients have developed from DNA repair impairment. To explore the dynamic associations of *EGFR* mutation within molecular pathways, we performed probabilistic analysis to identify genes that were highly co-mutated. Interestingly, ES NSCLCs demonstrated profound levels of genes that were highly co-mutated, including *PIK3CA* and *BRCA2*, whereas AS tumors showed significant co-enrichment of *EGFR* with *ROS1* and *ATRX* (*Figure 1E*). Moreover, *TP53* mutation showed mutual exclusivity with several genes, including *CTNN1B*. These results collectively suggest that ES and AS *EGFR*-mutant NSCLCs are defined by extensive molecular diversity and entail different treatment approaches.

Major canonical pathway alteration frequencies between ES and AS NSCLCs

As several core oncogenic pathways are frequently dysregulated in lung cancer, we aim to determine whether there were considerable levels of molecular disparity at the pathway level between ES and AS NSCLCs. We evaluated six canonical signaling pathways with frequent genetic alterations and discovered that RTK-RAS, PI3K, cell cycle, and p53 did not demonstrate much difference between the two groups. On the contrary, AS tumors showed enrichment of mutations in the NOTCH and histone modification pathways via mutations in *NCOR1*, *SPEN*, and *SMARCA4*, respectively (*Figure 2A*). Furthermore, despite the similar levels of dysregulations in the RTK-RAS-PI3K pathway, AS tumors showed profound levels of somatic mutation in *MTOR*, whereas ES tumors showed recurrent mutations in *ERBB2*. These results collectively suggest that ES and AS NSCLCs may have propagated through different underlying mechanisms that define their unique molecular characteristics and properties.

Clinical actionability of ES and AS NSCLCs reveals enrichments of PIK3CA 542K mutations

As most *EGFR*-mutant NSCLCs invariably acquire treatment resistance to EGFR-TKIs, we seek to identify alternative targets that are therapeutically exploitable. We leveraged the OncoKB knowledge database to systematically identify the potential target of choice

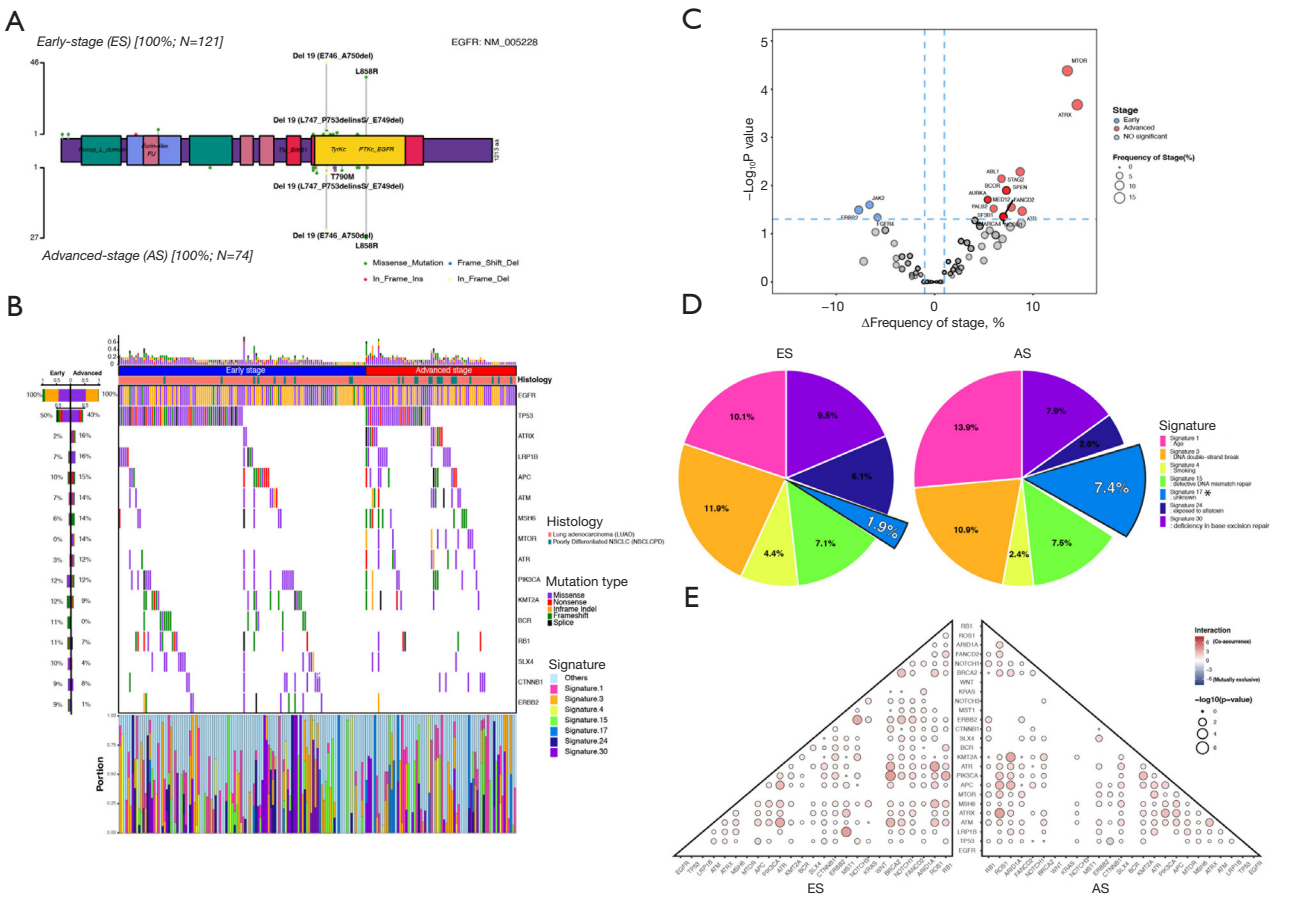

Figure 1 Genomic landscape of the K-MASTER *EGFR*-mutant NSCLC cohort. (A) Lollipop plot showing the distribution of *EGFR* mutations according to disease stage. (B) Genomic landscape of major cancer driver mutations across samples. The top panel depicts the proportion of each mutation across tumor types. The upper middle panel indicates the disease stage of each sample (blue, ES; red, AS). The center middle panel shows histological subtypes. The left bar represents the percentage of tumors within the K-MASTER *EGFR*-mutant NSCLC cohort for each mutation type. The lower middle panel displays the distribution of samples according to mutation type. The bottom panel shows the proportion of mutational signatures in each sample. Mutational signatures 1, 3, 15, 17, 24, and 30 were included based on a contribution of ≥7% in either ES or AS tumors. Signature 4, which is associated with smoking, was additionally included. (C) Volcano plot showing differences in mutation frequency between ES and AS tumors (X-axis) and statistical significance (Y-axis). Mutations significantly enriched in ES tumors are shown in blue, whereas those enriched in AS tumors are shown in red. (D) Pie charts illustrating the distribution of mutational signatures within each disease stage. Percentages indicate the relative contribution of each signature (e.g., signature 1 accounts for 10.1% in ES tumors and 13.9% in AS tumors). *, statistical significance (P<0.05, Wilcoxon rank-sum test). (E) Genetic interaction analysis stratified by disease stage. Red indicates co-occurring mutations, whereas blue indicates mutually exclusive interactions. AS, advanced-stage; *EGFR*, epidermal growth factor receptor; ES, early-stage; NSCLC, non-small cell lung cancer.

between ES and AS tumors. As expected, *EGFR* mutations, including both Del19 and L858R, were the most obvious candidates. However, we further discovered that both ES and AS tumors harbored considerable levels of mutations in the *PIK3CA* gene. Somatic alterations of the *PIK3CA* gene have been shown to play an important role in the activation of the PI3K-AKT-mTOR signaling pathway and promoting cellular growth (18). *PIK3CA* gene consists of five major functional domains, including the p85 binding domain, RAS binding domain, C2, helical, and kinase domains (19). Interestingly, we discovered that mutations that lead to E542K amino acid substitution in the helical domain were predominantly observed only in ES tumors (*Figure 2B*,*2C*).

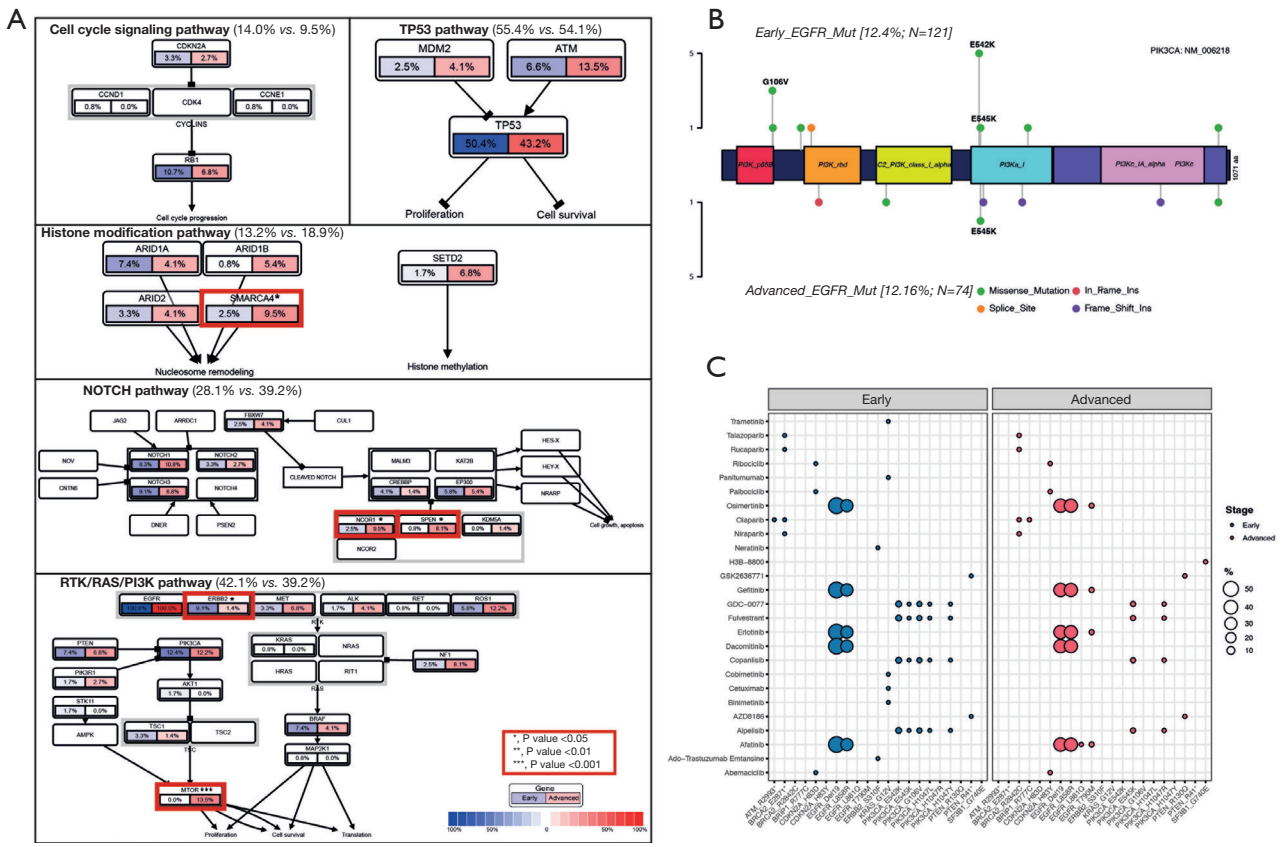

Figure 2 Pathway-level alteration frequencies between ES and AS tumors and stage-specific mutation patterns of *PIK3CA*. (A) Alteration frequencies of genes across major canonical pathways stratified by disease stage. Pathways with higher alteration frequencies in AS tumors are shown in red, whereas those enriched in ES tumors are shown in blue. Canonical pathway definitions were obtained from PathwayMapper (https://pathwaymapper.org). (B) Lollipop plot illustrating the distribution of *PIK3CA* mutations across the gene according to disease stage. (C) Stage-specific differences in potentially actionable alterations involving *PIK3CA*. Treatment options were annotated based on the OncoKB knowledge database. AS, advanced-stage; *EGFR*, epidermal growth factor receptor; ES, early-stage.

Molecular diversification of ES and AS NSCLCs based on major EGFR mutant subtypes

A substantial number of studies have coherently identified enrichments of *EGFR* activating mutations at various "hotspot" domains, including L858R mutation at exon 21 and in-frame deletion at exon 19, especially in never-smoker females in East Asian populations (8,15). In contrast to patients with Del19 mutant tumor, patients with L858R mutation often exhibited shorter survival probability as well as a higher likelihood to develop malignant pleural effusion and tumor invasion (8,20). Consistent with previous reports, *EGFR* Del19 was the most frequent genomic event in our cohort, followed by L858R mutation. Notably, the Del19 was more prevalent in ES tumors, whereas AS NSCLCs demonstrated similar levels of both Del19 and L858R mutations (*Figure 3A*). When we assessed the overall tumor mutational burden (TMB) in ES and AS NSCLCs, we discovered that L858R-mutant tumors were marked by a significantly higher rate of TMB compared to Del19 tumors only in the ES NSCLC patients, while AS patients did not show much difference (Figure S2A). AS NSCLC patients with Del19 exhibited more favorable clinical outcomes compared to other *EGFR* subtypes, while ES tumors did not demonstrate significant variations (*Figure 3B,3C*). Next, we stratified patients based on distinct *EGFR* mutant subtypes and consistently found that patients harboring Del19 demonstrated the most prolonged survival probability compared to others (Figure S2B). A significant difference in terms of overall TMB between ES and AS NSCLCs was only observed in Del19 tumors, where AS tumors

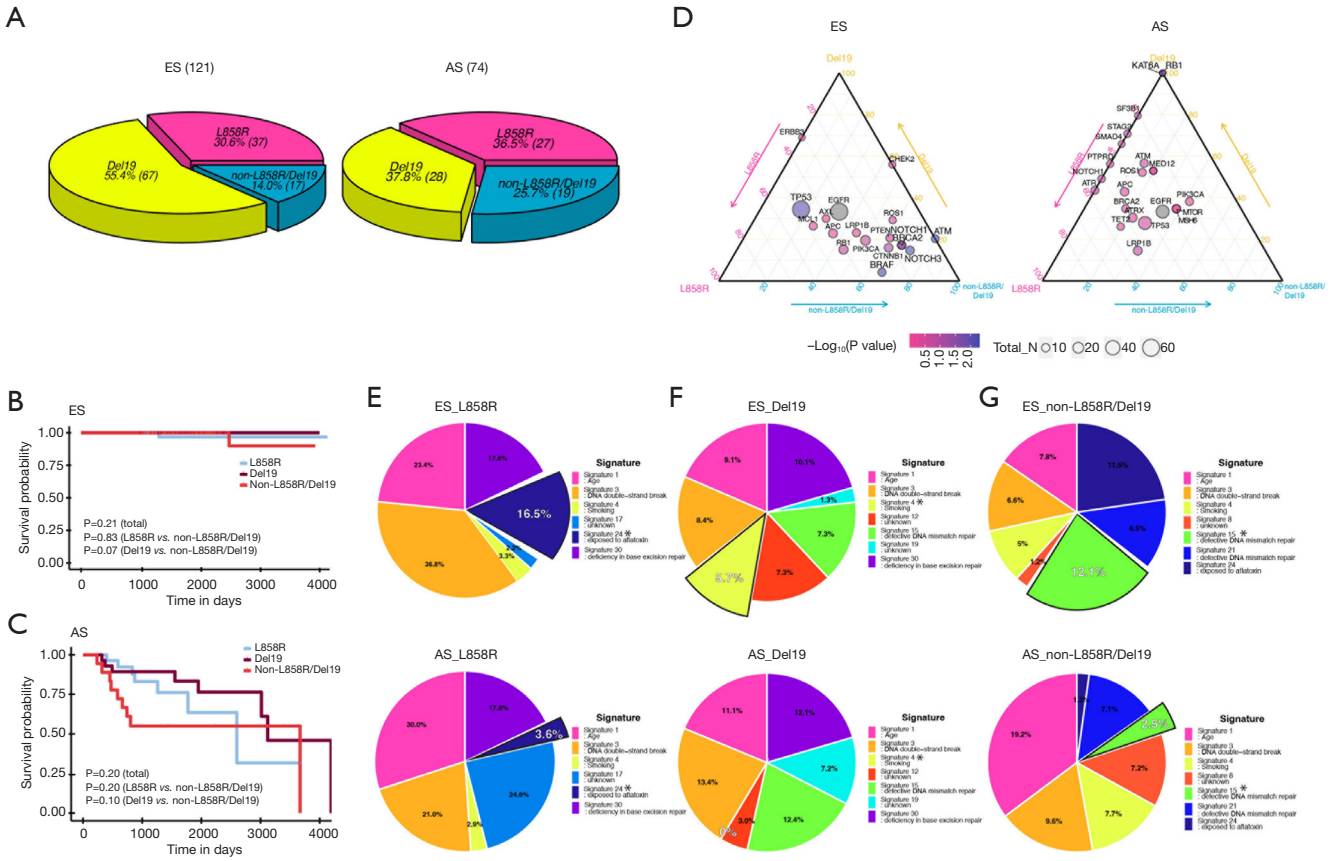

Figure 3 Comparison characteristic of subtypes in *EGFR* mutation NSCLC according to stage. (A) Distribution and relative frequencies of *EGFR* mutation subtypes in ES and AS tumors. (B) Kaplan-Meier estimates of OS in ES patients stratified by *EGFR* mutation subtype. P values were calculated using the log-rank test. (C) Kaplan-Meier estimates of OS in AS patients stratified by *EGFR* mutation subtype. P values were calculated using the log-rank test. (D) Ternary diagram illustrating mutation frequencies across L858R, Del19, and non-L858R/Del19 *EGFR* mutation groups. The size of each node represents the total number of tumors harboring the corresponding mutation in the *EGFR*-mutant NSCLC cohort, and the color scale indicates the significance of relative mutation frequencies. (E-G) Pie charts showing the distribution of mutational signatures according to *EGFR* mutation subtype. (E) Signatures observed in tumors with L858R mutations include signatures 1, 3, 4, 17, 24, and 30. (F) Signatures observed in Del19-mutant tumors include signatures 1, 3, 4, 12, 15, 19, and 30. (G) Signatures observed in tumors with non-L858R/Del19 mutations include signatures 1, 3, 4, 8, 15, 21, and 24. *, statistical significance (P<0.05, Wilcoxon rank-sum test). Only mutational signatures contributing more than 7% of total mutations are shown. AS, advanced-stage; Del19, exon 19 deletions; *EGFR*, epidermal growth factor receptor; ES, early-stage; NSCLC, non-small cell lung cancer; OS, overall survival.

showed higher levels of TMB (Figure S2C). When we investigated the frequency of other major driver mutations based on *EGFR* mutant subtypes, we discovered that *TP53* mutation was predominantly found in L858R-mutant tumors, whereas NOTCH pathway encoding molecules, including *NOTCH1* and *NOTCH3*, were highly occupied in non-L858R/Del19 tumors in ES NSCLCs (*Figure 3D*). Conversely, AS tumors exhibited considerable levels of heterogeneity, where major driver mutations were largely detected in L858R and Del19 mutant tumors. When we compared the mutational signature activities of ES and AS tumors, we discovered that all three *EGFR* subtype tumors constituted diverse mutational signature compositions where L858R-mutant tumors were characterized by enrichment of age- and HRD-associated signatures, while both Del19 and non-L858R/Del19 type tumors showed signs of DNA mismatch repair deficiency (MMRd) (*Figure 3E-3G*). Other prominent mutational signatures that were discovered in ES NSCLSs composed of aflatoxin, smoking, and MMRd-associated activities in L858R, Del19,

and non-L858R/Del19 tumors, respectively.

Identification of molecular determinants of clinical response to EGFR-TKI therapy

Among 195 patients, 84 received first- or second-generation EGFR-TKIs (gefitinib, erlotinib, dacomitinib, and afatinib), and 12 patients were treated with the third-generation inhibitor, osimertinib, as first-line therapy (*Figure 4A*). Tumor response was assessed using the Response Evaluation Criteria in Solid Tumors (RECIST) 1.1 criteria, and patients were classified as responders [complete response (CR), n=4; partial response (PR), n=23] or non-responders [stable disease (SD), n=16; progressive disease (PD), n=9] based on overall response rate (ORR). As expected, responder patients demonstrated significantly prolonged progression-free survival (PFS) compared to non-responders (median PFS: 486 *vs.* 167 days; Figure S3A). When we assessed the distribution of major *EGFR* mutation subtypes, L858R mutations were more prevalent in responders, whereas non-responders more frequently harbored T790M or non-hotspot *EGFR* mutations (Figure S3B). Stratification according to PFS using a 6-month threshold further highlighted our findings, where patients with longer PFS (>6 months) were enriched for L858R mutations, while those with shorter PFS exhibited a higher incidence of T790M mutations.

Among other major driver mutations, *TP53* and *NOTCH1* were distributed similarly between responder and non-responder groups. However, mutations in *MTOR* and *ATM* were more common in responder patients, while *KIT* and *ALK* point mutations were predominantly enriched in non-responders (*Figure 4B*). Notably, when we assessed the clinical impact of individual somatic mutations, we discovered that mutations in *ALK*, *SMO*, and *FANCA* were associated with unfavorable prognosis. Conversely, *EP300*, *PIK3R1*, *ERBB2*, and *RICTOR* mutations were associated with improved survival outcomes (*Figure 4C*). EP300 is a transcriptional co-activator and histone acetyltransferase that has been implicated in diverse oncogenic and tumor-suppressive processes, depending on cellular context. While EP300 has been reported to promote tumor progression in certain malignancies, emerging evidence suggests that EP300 alterations may also exert context-dependent effects on tumor behavior and treatment response (21). Consistent with prior studies reporting co-occurrence of *EGFR* and *ALK* alterations in a subset of NSCLC patients, ALK-mutated tumors in our cohort demonstrated poor response to EGFR-TKIs, supporting a potential resistance mechanism. To further elucidate the prognostic implications of these alterations, we conducted hazard ratio analyses of significantly mutated genes. Consistently, both *ALK* and *FANCA* mutations were associated with elevated risk, while *EP300* and *PIK3R1* mutations correlated with reduced risk, aligning with their respective distributions in favorable prognosis groups (*Figure 4D*,*4E*). Collectively, our results highlighted those patients with *EGFR* L858R mutations and co-occurring mutations in *EP300* and *PIK3R1* demonstrated favorable clinical response to EGFR-TKIs, whereas patients with *EGFR* T790M, *ALK*, or *FANCA* mutations exhibited increased treatment resistance and poor clinical outcomes. To provide additional clinical context regarding treatment heterogeneity, PFS was descriptively compared across individual EGFR-TKI agents. Interestingly, patients who received afatinib demonstrated the longest PFS, followed by osmirtinib, gefitinib, and erlotinib (Figure S4).

Discussion

With the global expansion of precision oncology programs, molecular characterization of tumors has become a fundamental component of cancer management, particularly for patients with limited therapeutic options following failure of standard-of-care treatments. We previously demonstrated the clinical feasibility of prospective genomic sequencing to investigate complex genomic landscapes across various solid tumors and to assess molecular heterogeneity and clinical utility in different genetic and clinical contexts. Building on this work, the present study comprehensively analyzed the molecular characteristics of 195 patients with NSCLC across different disease stages.

Lung cancer remains one of the leading causes of cancer-related mortality worldwide, with a particularly high burden in East Asian populations. Although chronic tobacco smoking accounts for a large proportion of NSCLC cases, a substantial subset of patients—predominantly female individuals of Asian ancestry—develop NSCLC without a history of smoking. In this population, activating mutations in the EGFR gene, particularly Del19 and the exon 21 L858R point mutation, are frequently observed. Numerous experimental and clinical studies have established the oncogenic roles of these alterations, and the development of EGFR-TKIs has dramatically improved outcomes for patients with *EGFR*-mutant NSCLC. However, because surgical resection remains the standard first-line treatment for ES NSCLC, the biological and clinical implications

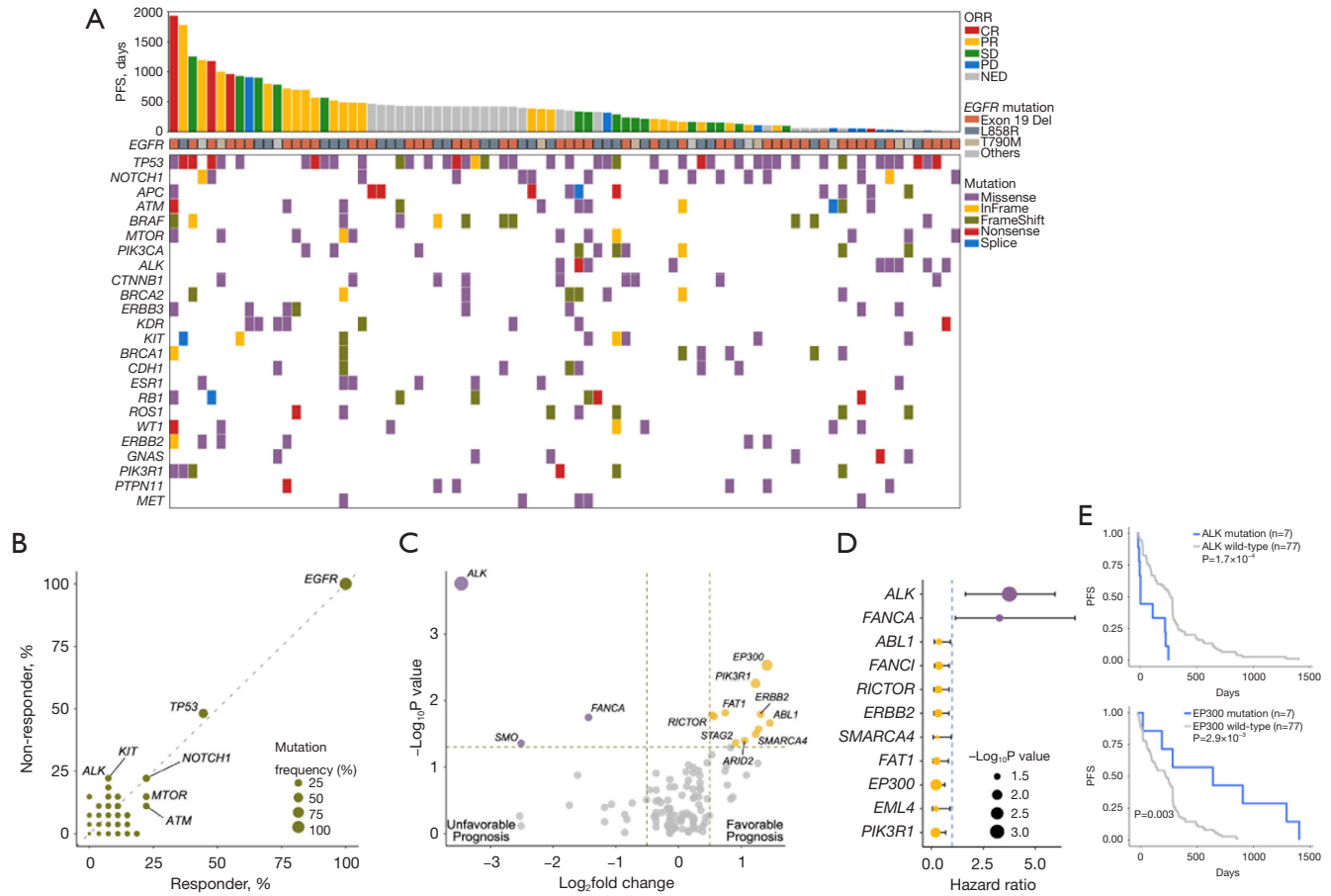

Figure 4 Clinical response to EGFR-TKI treatments relates specific genes. (A) Genomic landscape of major cancer driver mutations ordered by PFS. The top panel displays PFS and ORR for each sample. The middle panel indicates the *EGFR* mutation subtype, and the bottom panel shows the distribution of mutations across samples. (B) Comparison of mutation frequencies between non-responder (Y-axis) and responder (X-axis) cohorts. The size of each node represents the mutation frequency (%). (C) Volcano plot showing differences in mutation frequency between tumors with favorable and unfavorable prognosis following EGFR-TKI therapy (X-axis) and statistical significance (Y-axis). Mutations significantly enriched in the favorable prognosis group are shown in yellow, whereas those enriched in the unfavorable prognosis group are shown in purple. Prognosis groups were defined based on survival comparisons between patients harboring the mutation and those with wild-type status. If the patients harboring the corresponding mutation demonstrated increased survival, then the group is annotated as "favorable" and vice versa for the "unfavorable" group. (D) Forest plot illustrating hazard ratios for individual gene mutations. Circles represent hazard ratio estimates, and horizontal bars indicate the 95% confidence intervals. (E) Kaplan-Meier curves for PFS stratified by *ALK* mutation status (upper) and *EP300* mutation status (lower). CR, complete response; EGFR, epidermal growth factor receptor; NED, no evidence of disease; ORR, overall response rate; PD, progressive disease; PFS, progression-free survival; PR, partial response; SD, stable disease; TKI, tyrosine kinase inhibitor.

of *EGFR* alterations in ES disease remain incompletely understood.

In this study, we performed comprehensive genomic profiling of 121 ES and 74 AS *EGFR*-mutant NSCLC patients to delineate stage-specific molecular features. *EGFR* mutations were predominantly located at two hotspot regions—Del19 and exon 21 L858R mutations—with Del19 mutations occurring more frequently in ES tumors. *TP53* alterations represented the most common co-occurring driver mutation in both groups. In contrast, AS tumors were enriched for mutations in *MTOR*, *ATRX*, *STAG2*, *ABL1*, and *SPEN*, whereas ES tumors more frequently

harbored mutations in *JAK2*, *ERBB2*, and *FGFR4*. Pathway-level analyses further revealed a higher degree of genetic deregulation involving NOTCH signaling and histone modification pathways in AS tumors compared with ES tumors.

We further observed molecular heterogeneity across disease stages within specific EGFR mutation subtypes. AS tumors harboring Del19 mutations exhibited significantly higher TMB, whereas ES tumors with uncommon EGFR mutations demonstrated enrichment of DNA repair-related alterations. Using the OncoKB knowledge base, we explored potentially actionable alterations and identified enrichment of the *PIK3CA E542K* mutation in ES tumors, suggesting possible avenues for future translational investigation.

Several limitations of this study should be acknowledged. First, the EGFR-TKI treatment landscape represented in this cohort reflects historical clinical practice at the time of patient enrollment, during which third-generation EGFR-TKIs, including osimertinib and lazertinib, had not yet been approved or widely adopted as standard first-line therapy. Consequently, most patients received first- or second-generation EGFR-TKIs. Accordingly, the observed association between T790M mutation status and overall survival should be interpreted with caution, as it may be influenced by treatment availability and sequencing rather than solely reflecting the intrinsic biological impact of the mutation.

Second, although an exploratory comparison of PFS across individual EGFR-TKI agents was performed, detailed information regarding prior treatments and comorbidities was not consistently available. Treatment heterogeneity and unmeasured clinical confounders may therefore have influenced the observed genomic associations. Third, the limited number of patients harboring individual genetic alterations restricts statistical power, and mutation-specific findings should be regarded as exploratory, which warrants further investigation.

We additionally evaluated pharmacogenomic associations in a subset of 84 NSCLC patients who received EGFR-TKIs. Distinct mutational patterns were observed between patients with favorable and unfavorable responses to EGFR-TKI therapy. Poor responders were characterized by higher frequencies of *TP53*, *KIT*, and *ALK* point mutations, whereas favorable responders exhibited enrichment of *MTOR*, *ATM*, *EP300*, and *PIK3R1* mutations. Consistent with prior studies, *TP53* mutations were associated with inferior outcomes, supporting their role as negative predictive biomarkers of EGFR-TKI efficacy. *ALK* point mutations may reflect alternative oncogenic signaling pathways contributing to resistance in *EGFR*-mutant tumors. Conversely, enrichment of *EP300* and *PIK3R1* mutations in responders may indicate a less aggressive tumor phenotype or greater dependence on EGFR signaling; however, these interpretations remain speculative.

To further contextualize our findings, we performed a stage-based comparison within the TCGA NSCLC cohort (Figure S5). Although a broader distribution of EGFR mutation sites and a more diverse spectrum of co-occurring genetic alterations were observed in the TCGA dataset, this cohort is predominantly composed of ES tumors. As a result, ES-associated mutational patterns may be overrepresented, limiting direct comparability with our cohort. These differences are therefore more likely to reflect cohort composition rather than true biological inconsistency. Further experimental investigations will be required to determine their biological significance and potential clinical relevance. Future studies incorporating functional assays and larger independent validation cohorts are warranted to clarify the mechanistic contributions of these alterations and to explore their potential utility as therapeutic targets or biomarkers. Overall, while curative-intent local therapies and EGFR-TKI are the best treatment options for ES and AS patients, respectively, our results highlight genomic alterations that are associated with response to current standard treatment modalities.

Conclusions

The current study demonstrates that ES and AS *EGFR*-mutant NSCLC exhibit dynamic molecular disparity and differential genomic determinants of EGFR-TKI response. These findings underscore the importance of stage- and subtype-specific genomic profiling and provide a framework for future translational and biomarker-driven studies in *EGFR*-mutant NSCLC.

Acknowledgments

None.

Footnote

Reporting Checklist: The authors have completed the REMARK reporting checklist. Available at https://tlcr.amegroups.com/article/view/10.21037/tlcr-2025-1-1354/rc

Data Sharing Statement: Available at https://tlcr.amegroups.com/article/view/10.21037/tlcr-2025-1-1354/dss

Peer Review File: Available at https://tlcr.amegroups.com/article/view/10.21037/tlcr-2025-1-1354/prf

Funding: This research was supported by a grant from the Korea Health Technology R&D Project by the Korea Health Industry Development Institute (KHIDI) (No. HI22C0823). The K-MASTER project was supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (No. HI17C2206).

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://tlcr.amegroups.com/article/view/10.21037/tlcr-2025-1-1354/coif). J.H., J.K.S., and Y.J.C. are employees of oncoMASTER Inc. The other authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki and its subsequent amendments. The study was approved by the Ethics Committee of Korea University Anam Hospital (approval No. 2017AN0401). All participating hospitals were informed and agreed to the study. Written informed consent was obtained from all patients prior to sample collection.

References

1. Park KH, Choi JY, Lim AR, et al. Genomic Landscape and Clinical Utility in Korean Advanced Pan-Cancer Patients from Prospective Clinical Sequencing: K-MASTER Program. Cancer Discov 2022;12:938-48.
2. Morash M, Mitchell H, Beltran H, et al. The Role of Next-Generation Sequencing in Precision Medicine: A Review of Outcomes in Oncology. J Pers Med 2018;8:30.
3. Oliver GR, Hart SN, Klee EW. Bioinformatics for clinical next generation sequencing. Clin Chem 2015;61:124-35.
4. Devarakonda S, Morgensztern D, Govindan R. Genomic alterations in lung adenocarcinoma. Lancet Oncol 2015;16:e342-51.
5. Chen J, Yang H, Teo ASM, et al. Genomic landscape of lung adenocarcinoma in East Asians. Nat Genet 2020;52:177-86.
6. Li D, Huang Y, Cai L, et al. Genomic landscape of metastatic lung adenocarcinomas from large-scale clinical sequencing. Neoplasia 2021;23:1204-12.
7. Kour O, Garg M. The genomic landscape of lung adenocarcinoma—insights towards personalized medicine. Proc Indian Natl Sci Acad 2021;87:562-77.
8. Chen YJ, Roumeliotis TI, Chang YH, et al. Proteogenomics of Non-smoking Lung Cancer in East Asia Delineates Molecular Signatures of Pathogenesis and Progression. Cell 2020;182:226-244.e17.
9. Pi C, Xu CR, Zhang MF, et al. EGFR mutations in early-stage and advanced-stage lung adenocarcinoma: Analysis based on large-scale data from China. Thorac Cancer 2018;9:814-9.
10. Paez JG, Jänne PA, Lee JC, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science 2004;304:1497-500.
11. Herbst RS, Morgensztern D, Boshoff C. The biology and management of non-small cell lung cancer. Nature 2018;553:446-54.
12. Pao W, Miller V, Zakowski M, 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 2004;101:13306-11.
13. Chaft JE, Rimner A, Weder W, et al. Evolution of systemic therapy for stages I-III non-metastatic non-small-cell lung cancer. Nat Rev Clin Oncol 2021;18:547-57.
14. Goldstraw P, Chansky K, Crowley J, et al. The IASLC Lung Cancer Staging Project: Proposals for Revision of the TNM Stage Groupings in the Forthcoming (Eighth) Edition of the TNM Classification for Lung Cancer. J Thorac Oncol 2016;11:39-51.
15. Yang CY, Yang JC, Yang PC. Precision Management of Advanced Non-Small Cell Lung Cancer. Annu Rev Med 2020;71:117-36.

16. Cui A, Hua H, Shao T, et al. Aflatoxin B1 induces Src phosphorylation and stimulates lung cancer cell migration. Tumour Biol 2015;36:6507-13.
17. Guo G, Li G, Liu Y, et al. Next-Generation Sequencing Reveals High Uncommon EGFR Mutations and Tumour Mutation Burden in a Subgroup of Lung Cancer Patients. Front Oncol 2021;11:621422.
18. Ludovini V, Bianconi F, Pistola L, et al. Phosphoinositide-3-kinase catalytic alpha and KRAS mutations are important predictors of resistance to therapy with epidermal growth factor receptor tyrosine kinase inhibitors in patients with advanced non-small cell lung cancer. J Thorac Oncol 2011;6:707-15.
19. Samuels Y, Velculescu VE. Oncogenic mutations of PIK3CA in human cancers. Cell Cycle 2004;3:1221-4.
20. Kelly WJ, Shah NJ, Subramaniam DS. Management of Brain Metastases in Epidermal Growth Factor Receptor Mutant Non-Small-Cell Lung Cancer. Front Oncol 2018;8:208.
21. Gronkowska K, Robaszkiewicz A. Genetic dysregulation of EP300 in cancers in light of cancer epigenome control-targeting of p300-proficient and -deficient cancers. Mol Ther Oncol 2024;32:200871.

Cite this article as: Yoon D, Lee JW, Cho BC, Kang EJ, Kim JS, Lim T, Yi SY, Kim YJ, Ahn MS, Kim YS, Park JH, Lim S, Park HS, Cho JH, Jang B, Lee JY, Kim J, Hong J, Koo H, Chung S, Shin SW, Kim YH, Sa JK, Choi YJ. Molecular and clinical disparity of *EGFR*-mutant non-small cell lung cancer (NSCLC) based on histopathological stage and *EGFR* molecular subtypes. Transl Lung Cancer Res 2026;15(3):50. doi: 10.21037/tlcr-2025-1-1354

