Integrating clinical outcomes with molecular simulations to guide the selection of non-classical EGFR-TKIs: A comprehensive analysis and development of treatment strategies

FANG WEI[1*], NIU NIU[2*], CHANG LI[3], GE GAO[3], NAN LI[3] and YAN WANG[1]

[1]Department of Medical Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, P.R. China; [2]Department of Medical Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital and Shenzhen Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Shenzhen, Guangdong 518116, P.R. China; [3]School of Computer and Information, Hefei University of Technology, Hefei, Anhui 230009, P.R. China

Received February 13, 2026; Accepted June 15, 2026

DOI: 10.3892/etm.2026.13272

Abstract. Non-classical EGFR mutations in non-small cell lung cancer (NSCLC) are uncommon and heterogeneous, and the optimal EGFR-tyrosine kinase inhibitor (TKI) strategy remains uncertain. This retrospective study analyzed 31 patients with advanced/metastatic, recurrent or post-resection NSCLC harboring non-classical EGFR mutations who were treated at the Chinese Academy of Medical Sciences Cancer Hospital (Beijing, China) between 2013 and 2020, with follow-up updated to June 30, 2024. EGFR mutation status was obtained from routine clinical molecular pathology reports based on clinically validated targeted next-generation sequencing and/or amplification refractory mutation system PCR. Molecular docking and Prime Molecular Mechanics Generalized Born Surface Area (MM-GBSA) calculations were used to evaluate drug-mutant EGFR binding. Clinical outcomes varied by mutation subtype and EGFR-TKI regimen; relatively favorable progression-free survival (PFS) was observed with dacomitinib in L861Q and with afatinib- or dacomitinib-based treatment in selected S768I compound mutations. Exploratory Spearman analysis suggested that lower docking scores and more negative MM-GBSA ΔG_bind values were associated with longer PFS. These findings provide hypothesis-generating evidence that molecular simulation may help inform personalized EGFR-TKI selection for non-classical EGFR mutations, although validation in larger, consecutive and prospective multicenter cohorts is required.

Correspondence to: Professor Niu Niu, Department of Medical Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital & Shenzhen Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, 113 Baohe Avenue, Longgang, Shenzhen, Guangdong 518116, P.R. China
E-mail: niuniuhop@hotmail.com

Professor Yan Wang, Department of Medical Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, 17 South Panjiayuan, Chaoyang, Beijing 100021, P.R. China
E-mail: wangyanmedonco@outlook.com

*Contributed equally

Key words: EGFR-TKI, Non-classical EGFR mutations, molecular simulation, non-small cell lung cancer, personalized therapeutic strategies

Introduction

The development of EGFR tyrosine kinase inhibitors (EGFR-TKIs) has significantly transformed the treatment paradigm for non-small cell lung cancer (NSCLC), particularly for patients with activating EGFR mutations, providing new avenues for targeted and effective therapy (1). While the development of multiple generations of EGFR-TKIs has provided increasingly effective treatment options (2), optimizing treatment for patients with non-classical EGFR mutations is a major clinical challenge (3). These non-classical variants represent ~10-18% of all EGFR mutations (4) and show unique biological characteristics and distinct treatment response patterns compared to classical mutations (5). Their complexity stems from structural diversity, which contributes to variable responses across different generations of EGFR-TKIs (6). The G719X mutation family, which is associated mainly with S768I, poses unique challenges in the selection of treatment strategies (7). These mutations affect the ATP-binding pocket of the EGFR kinase domain in ways that differ from classical mutations (8), potentially altering drug-binding characteristics and treatment outcomes (9). Although these mutations have significant clinical importance, their low frequency has led to limited representation in major clinical trials (10). Non-classical EGFR mutations remain poorly managed, with a knowledge gap in clinical guidelines that generally rely on small studies or case series (11). However, recent advances in computational methods, including protein structure prediction and molecular dynamics simulations, have

expanded the knowledge of EGFR mutants (12). Molecular modeling and artificial intelligence improve the prediction of drug-target interactions and support clinical evidence (13). This study fills this void by combining data from molecular simulations with clinical data to guide evidence-based approaches to clinically informed selection and sequencing of therapy (14). The findings revealed the mechanisms of response and resistance, providing insights into the development of personalized and more efficacious therapeutic strategies (15).

Materials and methods

Methods of data collection and determination for EGFR mutation. This retrospective study included 31 patients with advanced/metastatic, recurrent or post-resection NSCLC harboring non-classical EGFR mutations, including G719X/G719A, L861Q, S768I and compound mutation patterns. Patients were selected from the institutional lung cancer cohort if they had a documented non-classical EGFR mutation, received EGFR-TKI-based systemic therapy and had evaluable clinical follow-up data. All patients were treated at the Chinese Academy of Medical Sciences (CAMS) Cancer Hospital (Beijing, China) between January 2013 and December 2020 and received EGFR-TKI-based systemic therapy. Progression-free survival (PFS) during first-line and second-line treatment was evaluated as PFS1 and PFS2, respectively. Progression patterns after first-line therapy were classified into three categories based on clinical and radiographic assessment: Slow progression (SP), defined as gradual asymptomatic multi-site progression; local progression (LP), defined as oligoprogression limited to one or a few sites; and clinical progression (CP), defined as rapid symptomatic systemic progression (16,17). Follow-up data were last updated on June 30, 2024.

EGFR mutation status was determined as part of routine clinical molecular diagnostics in certified clinical molecular pathology laboratories. Molecular testing was performed using formalin-fixed paraffin-embedded tumor tissue and/or cytology specimens obtained at diagnosis or recurrence. According to specimen availability and institutional clinical practice during the study period, clinically validated targeted next-generation sequencing (NGS) and/or amplification refractory mutation system polymerase chain reaction (ARMS-PCR) assays were used. For NGS-based testing, commercially available targeted panels were employed, including GeneseeqPrime™ (425-gene panel; Geneseeq Technology Inc.), OncoScreen Plus (520-gene panel; Burning Rock Biotech Ltd.) and Genecast Comprehensive (769-gene panel with MinerVa® bioinformatics platform; Genecast Biotechnology Co., Ltd.), as well as an institution-validated in-house 56-gene targeted panel (CAMS Cancer Hospital) (a laboratory-developed test validated and accredited for routine clinical use within the certified molecular pathology laboratory of the institution; as an in-house clinical assay, no separate peer-reviewed methodological publication is available); all NGS libraries were sequenced on the Illumina NovaSeq platform (Illumina, Inc.). For ARMS-PCR-based testing, the Human EGFR Gene Mutations Detection Kit (Super-ARMS; cat no. AD-00007; Amoy Diagnostics Co., Ltd.) was used. All commercial assays were performed in certified clinical molecular pathology laboratories according to the manufacturers' instructions and institutional standard operating procedures. Mutation subtype classification and related molecular findings were retrospectively extracted from the final clinical molecular pathology reports.

No *de novo* PCR or sequencing experiments were performed by the investigators for this study. Raw sequencing files (FASTQ, BAM and VCF formats), PCR primer sequences, polymerase specifications and laboratory-specific proprietary assay parameters were not generated by the investigators and are retained by the respective certified clinical molecular pathology laboratories under institutional data governance policies, consistent with the retrospective design of this study. Accordingly, read alignment parameters, genome assembly procedures and raw sequencing accession numbers are not applicable to the present retrospective analysis. The data reported in this paper have been deposited in the OMIX, China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences (https://ngdc.cncb.ac.cn/omix; accession no. OMIX017364); de-identified patient-level clinical data are additionally provided in Tables SI-SIII; the corresponding molecular simulation metrics are presented in Table IV.

Structural modeling of non-classical EGFR mutations. To examine the influence of uncommon EGFR mutations on protein conformation and drug responsiveness, six EGFR variants were selected: G719A, G719A+S768I, G719C+S768I, G719S+S768I, L861Q and V769L+S768I. These variants were selected because they represented the principal single or compound non-classical EGFR mutations observed in the clinical cohort and had sufficient corresponding treatment-event data for exploratory molecular-clinical comparison. The investigation focused on the domain targeted by EGFR small-molecule inhibitors (residues G696-G1022). The three-dimensional structures of these mutants were predicted using the AlphaFold 3 algorithm (Google DeepMind; AlphaFold Server; https://alphafoldserver.com) (18). For each mutant, the corresponding amino acid sequence was compiled and formatted based on the input requirements of AlphaFold 3. Subsequently, computational prediction was performed to obtain the structures of the six mutant proteins. The resulting protein structures were imported into Maestro 13.9 software (Schrodinger, LLC) for visualization. The 'Protein Preparation Workflow' module in Maestro 13.9 was used to preprocess the protein structures, including adding missing hydrogen atoms, completing missing side chains, removing free water molecules and optimizing hydrogen bonds.

Molecular docking and Molecular Mechanics Generalized Born Surface Area (MM-GBSA) calculations. The three-dimensional structure files of the drugs selected for this study, including furmonertinib, icotinib, dacomitinib, erlotinib, gefitinib, osimertinib and afatinib, were retrieved and downloaded from the PubChem database (National Center for Biotechnology Information; https://pubchem.ncbi.nlm.nih.gov/). Using the LePro module in LeDock Win32 (http://www.lephar.com/software.htm), the protein PDB files were converted into a format compatible with LeDock, and corresponding dock files were generated within the input

directory. Molecular docking simulations between small molecules and proteins were then conducted using LeDock Win32 to determine the binding affinity scores of the protein-compound complexes. Following docking, the Prime MM-GBSA module in Maestro 13.9 (Schrodinger, LLC) was employed to compute the MM-GBSA binding free energies (ΔG_bind) for the ligand-protein binding conformations. Additionally, the 'Ligand Interaction Diagram' tool in Maestro 13.9 was utilized to construct two-dimensional representations of ligand-receptor interactions, visually illustrating the molecular binding features.

Evidence-based molecular simulation analysis of clinical outcomes. Spearman rank correlation analysis was carried out to explore possible relationships between PFS and molecular binding metrics (LeDock docking scores or MM-GBSA ΔG_bind as appropriate) from the results of *in silico* docking methods. Due to the small cohort, heterogeneous treatment regimens and expected high correlation between docking scores and MM-GBSA ΔG_bind, no multivariate linear regression model was used for inferring independent predictive effects; an ordinary least-squares multivariable linear regression was conducted solely as a collinearity diagnostic and confirmed high variance inflation factors (VIF=21.637), supporting the decision to restrict primary inference to individual Spearman correlations (19). The analysis was performed at the drug-mutation treatment event level for generating mechanistic hypotheses, rather than a validated patient-level predictive model. A two-sided $P<0.05$ was considered nominally significant and all results were interpreted as exploratory.

Results

Clinical outcomes of patients with non-classical EGFR-mutant NSCLC receiving EGFR-TKI treatment. The median PFS for patients carrying the G719A mutation was 7.4 months for first-line therapy and 10.3 months for second-line therapy. For individual first-line treatments, the median PFS values were as follows: Icotinib, 11.6 months; dacomitinib, 8.7 months; erlotinib, 7.7 months; gefitinib, 4.3 months; osimertinib, 4.0 months; and afatinib, 1.2 months (Table SI). The cohort comprised 23 female and 8 male patients (female:male ratio, 2.9:1); the median age was 63 years (range, 41-78 years), as summarized from the patient-level data in Tables SI-SIII.

The median PFS for patients with L861Q mutation was as follows: Median PFS for first-line therapy: 19.6 months; median PFS for second-line therapy: 9.1 months; median PFS for each first-line therapy: Dacomitinib: 24.9 months; afatinib: 15.7 months (Table SII).

The median PFS for patients with S768I compound mutation was as follows: Median PFS for first-line therapy: 11.3 months; median PFS for second-line therapy: 7.6 months. The median PFS for each first-line therapy was as follows: Dacomitinib: 23.1 months; afatinib-based therapy: 15.4 months; chemotherapy: 12.5 months; gefitinib: 8.8 months; and icotinib: 5.8 months (Tables SIII and I).

To analyze the first-line treatment efficacy for patients with specific EGFR mutations (G719A, S768I compound and L861Q), the median PFS and the number of patients who achieved a PFS exceeding 30 months were documented (Table II). The distribution of progressive disease types across different mutation subtypes is illustrated in Fig. 1. L861Q showed the longest median first-line PFS (19.6 months), while long-term PFS >30 months occurred in 1/8 G719A, 2/15 S768I compound and 1/8 L861Q cases. PD patterns were predominantly SP, occurring in 6/8 G719A, 5/8 L861Q and 7/15 S768I compound cases; LP occurred in 1/8 G719A and 2/15 S768I compound cases and CP occurred in 4/15 S768I compound cases.

Table I. Median first-line and second-line PFS (PFS1/PFS2) by S768I subtype mutation.

Gene mutation type	Median first-line PFS, months	Median second-line PFS, months
G719A+S768I	6.1 (3.2-17.1)	6.1 (5.4-23.9)
G719C+S768I	15.9 (8.8-20.6)	2.9 (1.0-12.2)
G719S+S768I	8.6 (4.4-30.2)	7.8 (7.6-12.5)
G719X+S768I	20.6 (5.3-44.4)	8.7 (3.0-10.7)
V769L+S768I	6.5	4.7

Values are expressed as the median (range); the range is not shown for the V769L+S768I subtype as it comprised a single patient. PFS, progression-free survival.

Table II. Information on long-term survivors.

Mutation	Median first-line PFS, months	Patients with PFS >30 months/total
G719A	7.4 (1.2-38.7)	1/8
S768I compound	11.3 (3.2-44.4)	2/15
L861Q	19.6 (9.3-41.8)	1/8

Values are expressed as the median first-line PFS (range) for each mutation group. PFS, progression-free survival.

Table III. Sequence alignment scores and structural deviations of non-classical EGFR mutations compared to those of WT-EGFR.

Gene mutation type	Alignment score	RMSD (Å)
G719A	0.042	1.023
L861Q	0.033	0.902
G719A+S768I	0.063	1.254
G719C+S768I	0.058	1.202
G719S+S768I	0.064	1.264
V769L+S768I	0.05	1.121

A smaller alignment score indicates better sequence alignment with the WT-EGFR. WT, wild-type; RMSD, root-mean-square deviation.

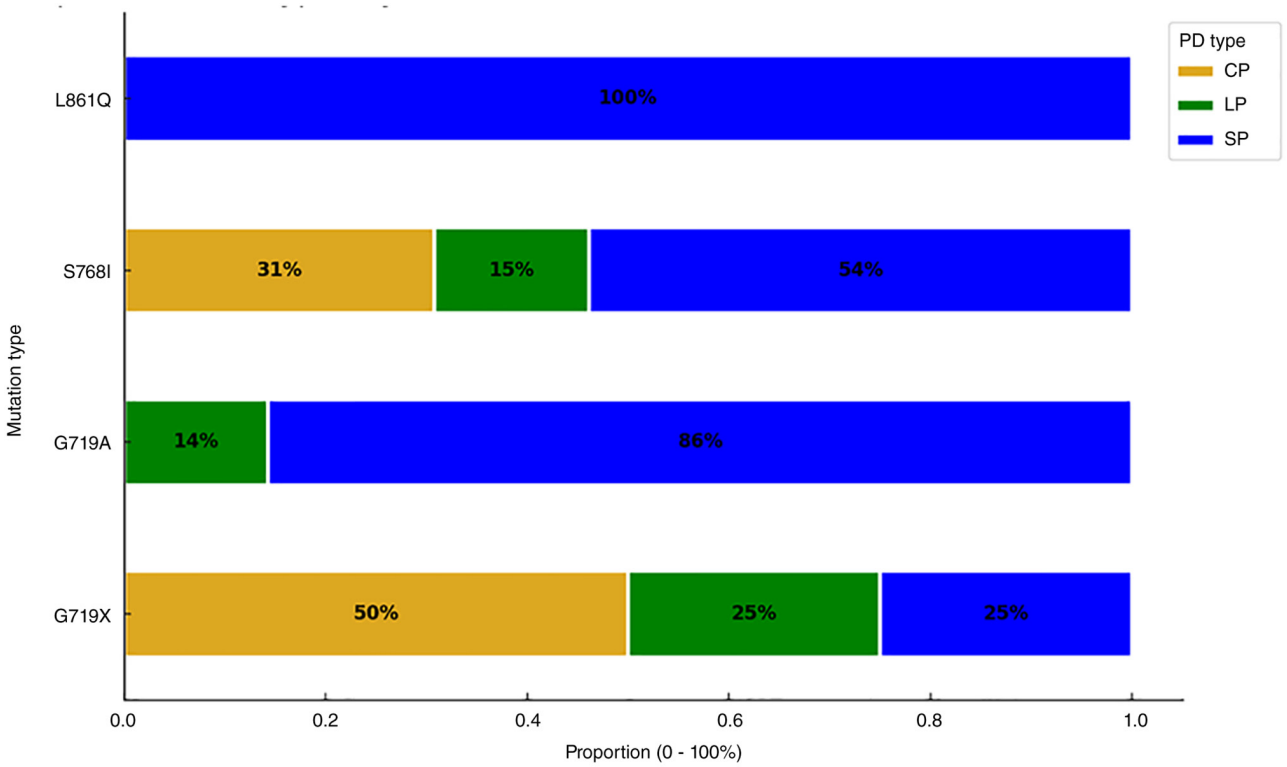

Figure 1. Proportion of PD types by different mutations. PD, progressive disease; CP, clinical progression; LP, local progression; SP, slow progression.

Structural effect of non-classical EGFR mutations. The alignment scores and root-mean-square deviation (RMSD) values for various non-classical EGFR mutations relative to the wild-type (WT) EGFR reference structure are presented in Table III. Alignment scores were generated using the multiple sequence viewer, where a lower score signifies better sequence alignment. The RMSD values [in angstroms (Å)] indicate the structural deviations of the mutated EGFR proteins from the WT protein. The mutations analyzed included single mutations (G719A, L861Q) and compound mutations involving S768I (G719A+S768I, G719C+S768I, G719S+S768I and V769L+S768I), highlighting the effect of these alterations on protein structure and drug-binding affinity. Structural superimposition of these mutant proteins with the WT EGFR is illustrated in Fig. 2. Overall, the single mutations G719A and L861Q showed smaller structural deviations from WT-EGFR (RMSD, 1.023 and 0.902 A, respectively) than most S768I compound mutations, among which G719S+S768I had the highest RMSD (1.264 A), suggesting greater conformational disruption in compound mutants.

Binding affinity of EGFR-TKIs for non-classical EGFR mutants. The relationships between different EGFR-TKIs, their binding affinities, as indicated by docking scores and MM-GBSA binding free energies (ΔG_bind) and the PFS of patients with specific non-classical EGFR mutations, are presented in Table IV. Lower docking scores and more negative MM-GBSA ΔG_bind values indicate stronger binding affinities between ligands and their corresponding EGFR mutant proteins. The table summarizes both single and compound mutations, illustrating the performance of different EGFR-TKIs under various genetic conditions.

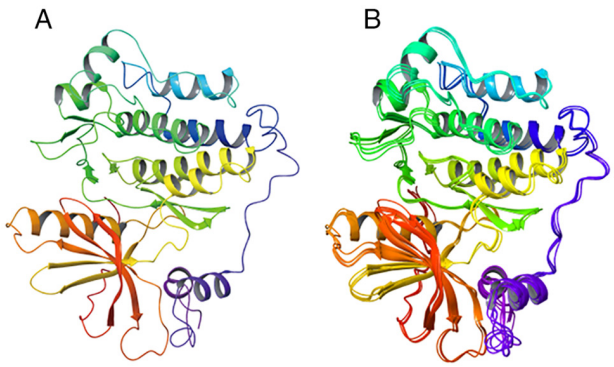

Figure 2. Structural superimposition of the WT-EGFR protein and six non-classical mutant EGFR proteins. (A) Three-dimensional structure of the wild-type EGFR protein; (B) Superimposed structures of the WT-EGFR protein and six non-classical mutant EGFR proteins. WT, wild-type.

To clarify the variations in binding interactions between EGFR-TKIs and WT or non-classical EGFR mutants, two-dimensional binding interaction diagrams shown in Figs. 3 and 4 were analyzed. Fig. 3 depicts the binding modes of several EGFR-TKIs with WT EGFR and two single-point mutants (G719A and L861Q). These single-site mutations lead to notable conformational rearrangements within the ATP-binding pocket, altering hydrogen bonding and hydrophobic interaction patterns relative to those in the WT protein. Such structural alterations are likely to influence both the binding affinity and inhibitory potency of EGFR-TKIs.

A total of four compound EGFR mutants: G719A+S768I, G719C+S768I, G719S+S768I and V769L+S768I, are presented in Fig. 4. The addition of S768I modifies the EGFR-TKI

Table IV. Associations between the binding affinity of EGFR-tyrosine kinase inhibitors and the clinical outcomes of patients with various non-classical EGFR mutations.

Ligand	PFS, months	Docking score	MM-GBSA ΔG_bind, kcal/mol	Gene mutation type
Icotinib	11.6	-8.279	-89.68	G719A
Dacomitinib	8.7	-6.715	-78.54	G719A
Erlotinib	7.7	-9.532	-63.91	G719A
Gefitinib	4.3	-5.261	-49.24	G719A
Osimertinib	4	-5.138	-48.01	G719A
Furmonertinib	-	-9.1	-47.97	G719A
Afatinib	1.2	-4.375	-42.69	G719A
Afatinib	17.1	-7.594	-72.18	G719A+S768I
Gefitinib	8.8	-6.548	-61.77	G719A+S768I
Dacomitinib	6.1	-5.946	-55.12	G719A+S768I
Icotinib	6.1	-5.529	-51.87	G719A+S768I
Furmonertinib	-	-7.962	-50.76	G719A+S768I
Osimertinib	-	-6.314	-49.64	G719A+S768I
Erlotinib	-	-5.128	-46.71	G719A+S768I
Dacomitinib	15.9	-8.517	-92.35	G719C+S768I
Furmonertinib	-	-8.07	-85.63	G719C+S768I
Gefitinib	8.8	-7.803	-83.49	G719C+S768I
Erlotinib	-	-8.676	-82.14	G719C+S768I
Afatinib	-	-7.832	-78.24	G719C+S768I
Icotinib	-	-4.642	-65.72	G719C+S768I
Osimertinib	-	-5.967	-64.47	G719C+S768I
Dacomitinib	30.2	-9.851	-95.73	G719S+S768I
Afatinib	11.3	-7.991	-85.02	G719S+S768I
Osimertinib	-	-5.551	-69.57	G719S+S768I
Erlotinib	-	-5.561	-65.67	G719S+S768I
Gefitinib	12.5	-6.953	-65.29	G719S+S768I
Icotinib	5.8	-6.107	-57.55	G719S+S768I
Furmonertinib	-	-4.285	-56.44	G719S+S768I
Dacomitinib	24.9	-8.926	-88.41	L861Q
Afatinib	15.7	-7.382	-70.96	L861Q
Icotinib	-	-7.634	-57.33	L861Q
Furmonertinib	-	-5.61	-53.46	L861Q
Gefitinib	-	-7.559	-50	L861Q
Osimertinib	-	-7.719	-45.64	L861Q
Erlotinib	-	-6.267	-39.92	L861Q
Erlotinib	-	-8.34	-85.64	V769L+S768I
Furmonertinib	-	-8.328	-79.24	V769L+S768I
Gefitinib	-	-8.164	-75.23	V769L+S768I
Icotinib	-	-7.336	-74.57	V769L+S768I
Dacomitinib	-	-5.979	-74.46	V769L+S768I
Osimertinib	-	-8.333	-63.52	V769L+S768I
Afatinib	6.5	-6.287	-58.36	V769L+S768I

Lower docking scores and more negative MM-GBSA ΔG_bind values indicate greater binding affinity between the ligand and the EGFR mutant protein. MM-GBSA, Molecular Mechanics Generalized Born Surface Area; PFS, progression-free survival.

binding landscape, inducing significant conformational shifts in the ATP-binding domain. These shifts enhance or alter interactions, including additional hydrogen bonds and hydrophobic contacts, potentially explaining the different clinical responses observed in patients with these mutations, as shown by the binding affinities and outcomes in Table IV. Quantitatively, dacomitinib showed favorable binding and clinical outcome signals in L861Q (docking score, -8.926;

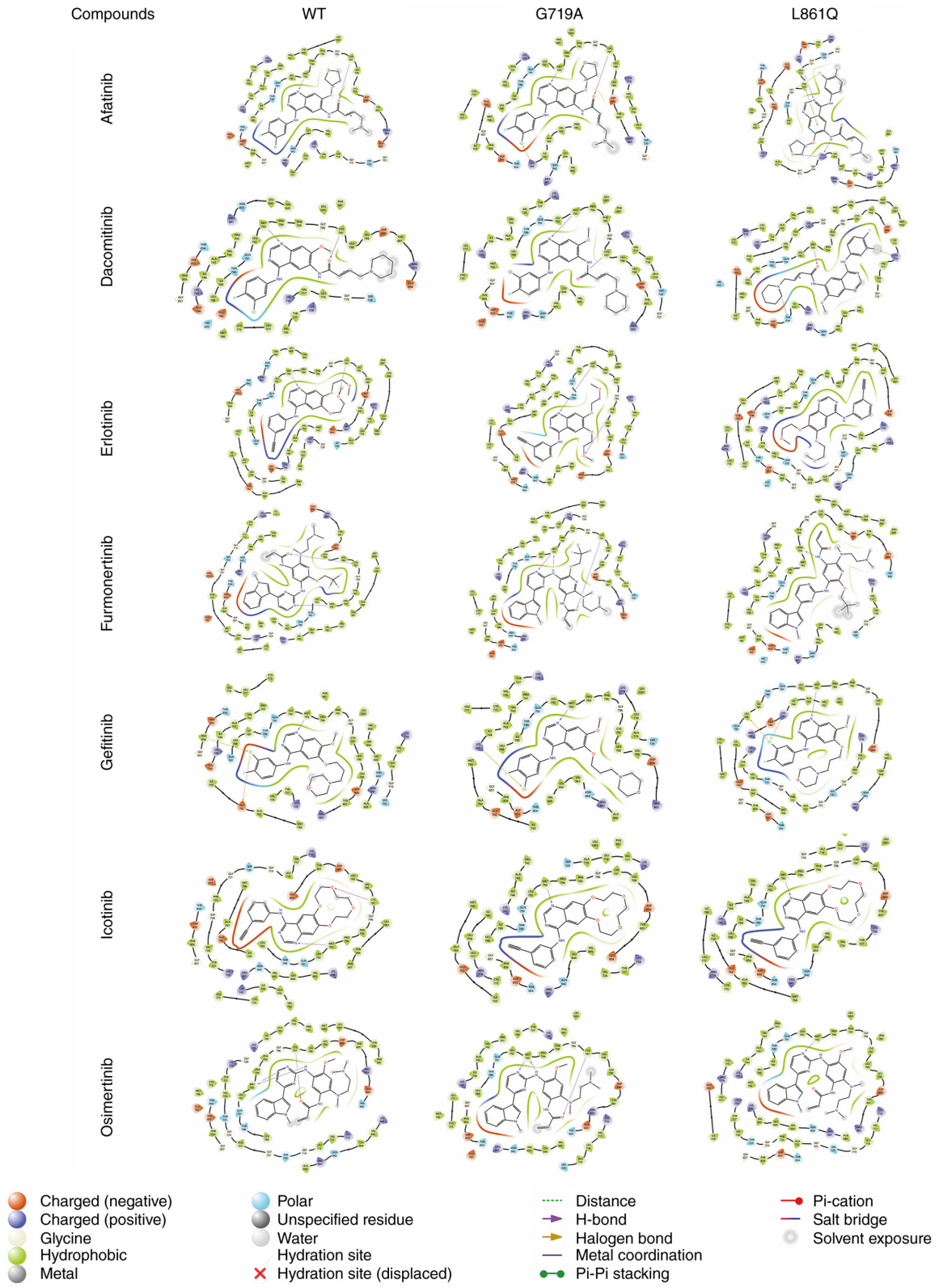

Figure 3. Two-dimensional binding interactions of EGFR-tyrosine kinase inhibitors with WT-EGFR and single-point mutants G719A and L861Q. WT, wild-type.

MM-GBSA ΔG_bind, -88.41 kcal/mol; PFS, 24.9 months) and G719S+S768I (docking score, -9.851; MM-GBSA ΔG_bind, -95.73 kcal/mol; PFS, 30.2 months), whereas afatinib in G719A+S768I showed a PFS of 17.1 months with an MM-GBSA ΔG_bind of -72.18 kcal/mol.

The binding interaction analyses depicted in Figs. 3 and 4 provide comprehensive insights into how specific non-classical EGFR mutations influence the molecular binding of EGFR-TKIs. Understanding this mechanism is crucial for optimizing the selection of EGFR-TKIs and developing

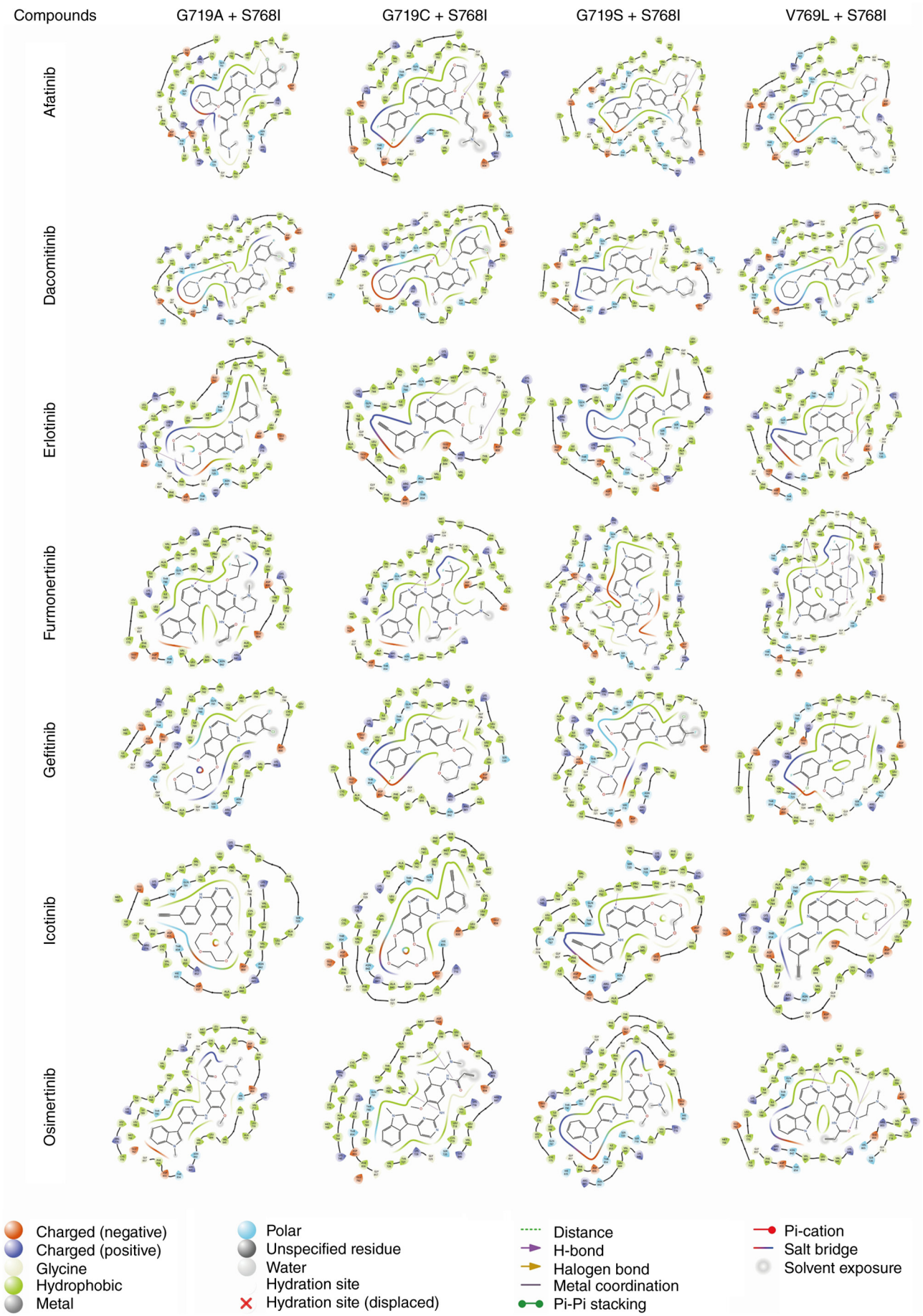

Figure 4. Two-dimensional binding interactions of EGFR- tyrosine kinase inhibitors with the compound EGFR mutants G719A+S768I, G719C+S768I, G719S+S768I and V769L+S768I.

Table V. Spearman correlation analysis between PFS, the docking score and MM-GBSA ΔG_bind.

Parameter	PFS, months	Docking score	MM-GBSA ΔG_bind, kcal/mol
PFS, months	1	-0.835 (P<0.0001)	-0.894 (P<0.0001)
Docking score	-0.835 (P<0.0001)	1	0.876 (P<0.0001)
MM-GBSA ΔG_bind, kcal/mol	-0.894 (P<0.0001)	0.876 (P<0.0001)	1

MM-GBSA, Molecular Mechanics Generalized Born Surface Area; PFS, progression-free survival.

Table VI. Multivariable linear regression analysis of progression-free survival.

Variable	Unstandardized coefficient (B)	Standard error	Standardized coefficient (beta)	t	P-value	VIF
Constant	-15.840	5.703	-	-2.778	0.013	-
Docking score	-4.369	3.442	-0.919	-1.269	0.222	21.637
MM-GBSA ΔG_bind, kcal/mol	0.061	0.318	0.140	0.194	0.849	21.637

Model statistics: R^2=0.612; adjusted R^2=0.564; F=12.644; overall P=0.001. Because docking score and MM-GBSA ΔG_bind were highly collinear, the regression coefficients should be interpreted cautiously. MM-GBSA, Molecular Mechanics Generalized Born Surface Area; VIF, variance inflation factor; Constant, regression intercept.

personalized treatment strategies for patients with NSCLC with non-classical EGFR mutations.

Exploratory correlation between molecular simulation results and clinical outcomes. The results of the Spearman correlation analysis are presented in Table V, based on the drug-mutation treatment events (first-line TKI, PFS, docking score and MM-GBSA ΔG_bind) presented in Table IV. Results of the collinearity diagnostic multivariate regression are provided in Table VI. PFS was negatively correlated with the docking score (r=-0.835, P<0.0001). Furthermore, PFS showed a strong negative correlation with MM-GBSA ΔG_bind (r=-0.894, P<0.0001), indicating that more negative values of MM-GBSA were correlated with better clinical outcomes. Furthermore, the docking score was strongly positively correlated with MM-GBSA ΔG_bind (r=0.876, P<0.0001), suggesting that both molecular simulation metrics were highly dependent on each other in this study. Thus, the molecular-clinical associations may represent exploratory and hypothesis-generating findings rather than definitive evidence of independent predictive effects. In Table VI, the constant represents the regression intercept rather than a biological predictor; docking score and MM-GBSA ΔG_bind were not independent predictors in this diagnostic model (P=0.222 and P=0.849, respectively), consistent with substantial collinearity (VIF=21.637).

Discussion

A comprehensive analysis of non-classical EGFR mutations may provide useful information for optimizing treatment strategies in a clinically challenging and underrepresented patient population. In the present study, integrating molecular simulation data with retrospective clinical outcomes provided a hypothesis-generating framework for EGFR-TKI selection for this patient subgroup, for which prospective trial evidence remains limited.

The present study revealed that dacomitinib exhibited stronger binding affinity to mutant proteins with the L861Q mutation compared to other EGFR-TKIs. Clinical data confirmed that first-line treatment with dacomitinib resulted in notably longer PFS, surpassing the PFS reported in previous studies with afatinib and osimertinib for L861Q mutations (20,21). The present research indicated that afatinib displayed favorable molecular binding parameters in patients with compound mutations, such as G719A+S768I, aligning with clinical observations of a PFS of 17.1 months, which is considerably longer than the treatment outcomes reported in other studies (21,22). Similarly, for compound mutations such as G719C+S768I or G719S+S768I, dacomitinib bound to mutant proteins with high efficacy. This finding was consistent with the clinical data. The present results highlighted the significant therapeutic variability among different structurally targeted drugs for atypical EGFR mutations (23-25). As these mutations are rare, pharmaceutical sponsors are often hesitant to invest in targeted clinical trials owing to an unfavorable cost-benefit ratio. As a result, clinical decisions typically rely on individual experiences of physicians, complicating the ability to offer patients the most accurate treatment options. The therapeutic efficacy of rare EGFR mutations has been predominantly assessed through retrospective analyses. While these studies support the present findings, including the therapeutic advantage of dacomitinib in managing uncommon compound mutations (20,26), retrospective designs are inherently limited by their lack of timeliness. In clinical practice, a prospective analytical framework is urgently required to generate reliable treatment guidance for the broad spectrum of uncommon EGFR mutations encountered in real time. The current study addresses this need by providing an

evidence-based and practical approach for selecting targeted therapies for rare EGFR mutations. This strategy helps bridge existing gaps in clinical research and facilitates the development of more precise and personalized treatment options for patients with atypical mutations. Patients with EGFR mutations typically exhibit LP or SP following targeted therapy, often retaining the original EGFR mutation (16). The clinical data of the present study confirmed the persistence of primary mutations during disease progression and validated drug sensitivity predictions for atypical mutations using molecular simulations. These predictions closely matched the clinical outcomes of second-line treatments, demonstrating the robustness of the model for both first-line and subsequent-line therapies. By accurately capturing drug-mutation interactions, the model provided a rational framework for selecting optimal treatments in cases with limited clinical data. The alignment between the predicted sensitivities and outcomes highlighted its ability to enhance precision medicine and improve PFS across multiple lines of therapy.

Owing to the heterogeneous and low-frequency occurrence of uncommon EGFR mutations (17,27,28), this study had several limitations. The retrospective nature of the study and the relatively small sample size of 31 patients may have limited the generalizability of the findings. As this was a single-center study, selection bias may have been unintentionally introduced and the diversity of the patient population may have been limited. The absence of a control group and the variability in treatment regimens further contributed to confounding factors, which may have influenced the results.

Another limitation relates to the retrospective ascertainment of EGFR mutation status. Molecular results were extracted from routine clinical pathology reports rather than generated *de novo* for this study. Consequently, raw sequencing files, PCR primer sequences and laboratory-specific proprietary assay parameters for NGS panels were not uniformly available to the investigators, although the ARMS-PCR kit catalogue number was retrievable from archived pathology records. In addition, the strong correlation between docking scores and MM-GBSA ΔG_bind indicates that these molecular metrics are not independent (19), and in high-correlation settings such collinearity may produce spurious predictor-outcome associations in regression models. For this reason, the molecular-clinical analysis was restricted to exploratory Spearman correlations and no multivariable regression was performed to infer independent predictive effects. Future prospective studies with standardized molecular testing, centrally reviewed raw genomic data and larger sample sizes are required to validate the proposed simulation-guided framework.

This is a valuable but limited strategy in light of unique therapeutic differences associated with non-classical EGFR mutations, such as exon 20 insertions, and other EGFR alterations, such as the overexpression of MET, which may affect the clinical outcomes of patients with NSCLC receiving EGFR-TKIs (21). As larger and prospective studies confirm these results, incorporating more biomarkers and combination therapies may help further personalize the approach to these patients. As machine learning and molecular dynamics simulation methods improve, predictions of drug-agent interactions can help researchers make better clinical decisions (12,15). Furthermore, investigating the tumor microenvironment and patient-specific factors, including comorbidities, will be pertinent. Translating these discoveries into clinical practice requires the close interaction of computational biologists and clinicians.

This study highlighted the power of molecular simulations in selecting EGFR-TKIs for treating NSCLC more effectively than traditional clinical analyses. The present findings may help improve patient stratification and therapeutic outcomes by tying molecular binding affinities to clinical endpoints. This approach can help address the existing knowledge gap and emphasize the necessity of personalized treatment strategies based on non-classical EGFR mutations.

In conclusion, these findings provide a proof-of-concept for mining molecular simulation data alongside retrospective clinical outcomes to guide the selection of EGFR-TKIs in patients harboring non-classical mutations. This study is exploratory in nature, but the heterogeneous treatment response observed suggests that mutation- and drug-specific molecular interpretations could be of potential value.

Despite being retrospective, single-center in scope and limited in sample size, differences in activity with EGFR-TKIs seen across non-classical mutation categories provide a rationale for further investigation. Larger multicenter datasets, standardized molecular diagnostics and prospective validation are necessary before molecular simulation-guided treatment recommendations can realistically be used in daily clinical practice.

Acknowledgements

Not applicable.

Funding

This work was supported by Shenzhen Science and Technology Program (grant no. JCYJ20220530153612028), the National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital & Shenzhen Hospital, CAMS & Peking Union Medical College (grant no. E010222007) and the Sanming Project of Medicine in Shenzhen (grant no. SZSM202211012).

Availability of data and materials

The sequencing data generated in the present study may be found in the OMIX database (China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences) under accession number OMIX017364 or at the following URL: https://ngdc.cncb.ac.cn/omix/release/OMIX017364. The other data generated in the present study may be requested from the corresponding author.

Authors' contributions

FW, NN and YW conceived and designed the study. FW, NN, CL, GG, NL and YW collected, analyzed and interpreted the data. FW and NN wrote the manuscript; NN, NL and YW provided critical revisions important for the intellectual content. NN and YW confirm the authenticity of all the raw data. All authors have read and approved the final version of the manuscript.

Ethics approval and consent to participate

This study was conducted in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of the National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, CAMS and Peking Union Medical College (approval no. 25/229-5175). Because of the retrospective nature of the study and the use of de-identified clinical data, the requirement for written informed consent was waived by the ethics committee.

Patient consent for publication

Not applicable.

Competing interests

All authors declare that they have no competing interests.

References

1. Cooper AJ, Sequist LV and Lin JJ: Third-generation EGFR and ALK inhibitors: Mechanisms of resistance and management. Nat Rev Clin Oncol 19: 499-514, 2022.
2. Ramalingam SS, Vansteenkiste J, Planchard D, Cho BC, Gray JE, Ohe Y, Zhou C, Reungwetwattana T, Cheng Y, Chewaskulyong B, *et al*: Overall survival with osimertinib in untreated, EGFR-Mutated advanced NSCLC. N Engl J Med 382: 41-50, 2020.
3. Robichaux JP, Le X, Vijayan RSK, Hicks JK, Heeke S, Elamin YY, Lin HY, Udagawa H, Skoulidis F, Tran H, *et al*: Structure-based classification predicts drug response in EGFR-mutant NSCLC. Nature 597: 732-737, 2021.
4. O'Kane GM, Bradbury PA, Feld R, Leighl NB, Liu G, Pisters KM, Kamel-Reid S, Tsao MS and Shepherd FA: Uncommon EGFR mutations in advanced non-small cell lung cancer. Lung Cancer 109: 137-144, 2017.
5. Frankell AM, Dietzen M, Al Bakir M, Lim EL, Karasaki T, Ward S, Veeriah S, Colliver E, Huebner A, Bunkum A, *et al*: The evolution of lung cancer and impact of subclonal selection in TRACERx. Nature 616: 525-533, 2023.
6. Yamaguchi O, Kasahara N, Soda H, Imai H, Naruse I, Yamaguchi H, Itai M, Taguchi K, Uchida M, Sunaga N, *et al*: Predictive significance of circulating tumor DNA against patients with T790M-positive EGFR-mutant NSCLC receiving osimertinib. Sci Rep 13: 20848, 2023.
7. Borgeaud M, Parikh K, Banna GL, Kim F, Olivier T, Le X and Addeo A: Unveiling the landscape of uncommon EGFR mutations in NSCLC-A systematic review. J Thorac Oncol 19: 973-983, 2024.
8. Jiang Y, Fang X, Xiang Y, Fang T, Liu J and Lu K: Afatinib for the treatment of NSCLC with uncommon EGFR mutations: A narrative review. Curr Oncol 30: 5337-5349, 2023.
9. Pizzutilo EG, Agostara AG, Oresti S, Signorelli D, Stabile S, Lauricella C, Motta V, Amatu A, Ruggieri L, Brambilla M, *et al*: Activity of osimeRTInib in non-small-cell lung Cancer with UNcommon epidermal growth factor receptor mutations: Retrospective Observational multicenter study (ARTICUNO). ESMO Open 9: 103592, 2024.
10. Wang P, Fabre E, Martin A, Chouahnia K, Benabadji A, Matton L and Duchemann B: Successful sequential tyrosine kinase inhibitors to overcome a rare compound of EGFR exon 18-18 and EGFR amplification: A case report. Front Oncol 12: 918855, 2022.
11. Ullas B, Shrinidhi N, Mansi S, Narayan S, Parveen J, Surender D, Joslia JT and Anurag M: All EGFR mutations are (not) created equal: focus on uncommon EGFR mutations. J Cancer Res Clin Oncol 149: 1541-1549, 2023.
12. Zou B, Lee VHF and Yan H: Prediction of sensitivity to gefitinib/erlotinib for EGFR mutations in NSCLC based on structural interaction fingerprints and multilinear principal component analysis. BMC Bioinformatics 19: 88, 2018.
13. Lococo F, Ghaly G, Chiappetta M, Flamini S, Evangelista J, Bria E, Stefani A, Vita E, Martino A, Boldrini L, *et al*: Implementation of artificial intelligence in personalized prognostic assessment of lung cancer: A narrative review. Cancers (Basel) 16: 1832, 2024.
14. He B, Dong D, She Y, Zhou C, Fang M, Zhu Y, Zhang H, Huang Z, Jiang T, Tian J and Chen C: Predicting response to immunotherapy in advanced non-small-cell lung cancer using tumor mutational burden radiomic biomarker. J Immunother Cancer 8: e000550, 2020.
15. Qureshi R, Basit SA, Shamsi JA, Fan X, Nawaz M, Yan H and Alam T: Machine learning based personalized drug response prediction for lung cancer patients. Sci Rep 12: 18935, 2022.
16. Harada D and Takigawa N: Oligoprogression in non-small cell lung cancer. Cancers (Basel) 13: 5823, 2021.
17. Jackman D, Pao W, Riely GJ, Engelman JA, Kris MG, Jänne PA, Lynch T, Johnson BE and Miller VA: Clinical definition of acquired resistance to epidermal growth factor receptor tyrosine kinase inhibitors in non-small-cell lung cancer. J Clin Oncol 28: 357-360, 2010.
18. Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O, Willmore L, Ballard AJ, Bambrick J, *et al*: Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature 630: 493-500, 2024.
19. Leeuwenberg AM, van Smeden M, Langendijk JA, van der Schaaf A, Mauer ME, Moons KGM, Reitsma JB and Schuit E: Performance of binary prediction models in high-correlation low-dimensional settings: A comparison of methods. Diagn Progn Res 6: 1, 2022.
20. Yang JC, Schuler M, Popat S, Miura S, Heeke S, Park K, Märten A and Kim ES: Afatinib for the treatment of NSCLC harboring uncommon EGFR mutations: A database of 693 cases. J Thorac Oncol 15: 803-815, 2020.
21. Wang C, Zhao K, Hu S, Dong W, Gong Y, Li M and Xie C: Clinical outcomes of gefitinib and erlotinib in patients with NSCLC harboring uncommon EGFR mutations: A pooled analysis of 438 patients. Lung Cancer 172: 86-93, 2022.
22. Wu YL, Cheng Y, Zhou X, Lee KH, Nakagawa K, Niho S, Tsuji F, Linke R, Rosell R, Corral J, *et al*: Dacomitinib versus gefitinib as first-line treatment for patients with EGFR-mutation-positive non-small-cell lung cancer (ARCHER 1050): A randomised, open-label, phase 3 trial. Lancet Oncol 18: 1454-1466, 2017.
23. Wu YL, Zhou C, Hu CP, Feng J, Lu S, Huang Y, Li W, Hou M, Shi JH, Lee KY, *et al*: Afatinib versus cisplatin plus gemcitabine for first-line treatment of Asian patients with advanced non-small-cell lung cancer harbouring EGFR mutations (LUX-Lung 6): An open-label, randomised phase 3 trial. Lancet Oncol 15: 213-222, 2014.
24. Oxnard GR, Lo PC, Nishino M, Dahlberg SE, Lindeman NI, Butaney M, Jackman DM, Johnson BE and Jänne PA: Natural history and molecular characteristics of lung cancers harboring EGFR exon 20 insertions. J Thorac Oncol 8: 179-184, 2013.
25. Arcila ME, Chaft JE, Nafa K, Roy-Chowdhuri S, Lau C, Zaidinski M, Paik PK, Zakowski MF, Kris MG and Ladanyi M: Prevalence, clinicopathologic associations, and molecular spectrum of ERBB2 (HER2) tyrosine kinase mutations in lung adenocarcinomas. Clin Cancer Res 18: 4910-4918, 2012.
26. Li HS, Yang GJ, Cai Y, Li JL, Xu HY, Zhang T, Zhou LQ, Wang YY, Wang JL, Hu XS, *et al*: Dacomitinib for advanced non-small cell lung cancer patients harboring major uncommon EGFR alterations: A dual-center, single-arm, ambispective cohort study in China. Front Pharmacol 13: 919652, 2022.
27. Attili I, Passaro A, Pisapia P, Malapelle U and de Marinis F: Uncommon EGFR compound mutations in non-small cell lung cancer (NSCLC): A systematic review of available evidence. Curr Oncol 29: 255-266, 2022.
28. Russo A, Franchina T, Ricciardi G, Battaglia A, Picciotto M and Adamo V: Heterogeneous responses to epidermal growth factor receptor (EGFR) tyrosine kinase inhibitors (TKIs) in patients with uncommon EGFR mutations: New insights and future perspectives in this complex clinical scenario. Int J Mol Sci 20: 1431, 2019.