Am J Cancer Res 2020;10(11):4005-4015
www.ajcr.us /ISSN:2156-6976/ajcr0119313

Original Article

Acquired rare recurrent *EGFR* mutations as mechanisms of resistance to Osimertinib in lung cancer and *in silico* structural modelling

Lin Lin[2*], Qiang Lu[3*], Ran Cao[4], Qiuxiang Ou[4], Yutong Ma[4], Hua Bao[4], Xue Wu[4], Yang Shao[5,6], Zhaoxia Wang[7], Bo Shen[1]

[1]Department of Oncology, Jiangsu Cancer Hospital, Jiangsu Institute of Cancer Research, Nanjing Medical University Affiliated Cancer Hospital, Nanjing, Jiangsu, China; [2]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, China; [3]Department of Thoracic Surgery, Tangdu Hospital, The Air Force Military Medical University, Xi'an, Shanxi, China; [4]Translational Medicine Research Institute, Geneseeq Technology Inc., Toronto, ON, Canada; [5]Nanjing Geneseeq Technology Inc., Nanjing, Jiangsu, China; [6]School of Public Health, Nanjing Medical University, Nanjing, Jiangsu, China; [7]Department of Oncology, The Second Affiliated Hospital of Nanjing Medical University, Nanjing, Jiangsu, China. [*]Equal contributors.

Received July 31, 2020; Accepted October 1, 2020; Epub November 1, 2020; Published November 15, 2020

Abstract: A growing number of progression on Osimertinib among *EGFR*-mutated lung cancers represents a great challenge clinically. Our study aims to gain insights into novel mechanisms of acquired resistance to Osimertinib. We performed genomic studies on 2 large independent cohorts of lung cancer patients with progressed diseases on different tyrosine kinase inhibitors (TKIs). *In silico* modeling was used to study the structural mechanism of selected EGFR mutations. Compared with the 1st-TKIs-resistant group, *EGFR* mutations C797S/G, L718Q/V, L792F/H were significantly more enriched in the Osimertinib-resistant cohort, whose sensitivities to Osimertinib were successfully predicted. Importantly, a total of 14 low-frequency *EGFR* mutations were exclusively or significantly observed in the Osimertinib-resistant group, 7 were predicted to dramatically reduce the binding affinity of EGFR to Osimertinib (G796S, V802F, T725M, Q791L/H, P794S/R). Analysis of pre-Osimertinib treatment samples of two patients supported that *EGFR* V802F and G796S were acquired during the treatment. In addition, *EGFR* G796S was predicted to be susceptible to gefitinib. This study represented the largest real-world data so far investigating Osimertinib resistance in *EGFR*-mutated lung cancer. We identified a collection of coexistent EGFR rare mutations and provided possible guidance for those patients who progressed on the first-line treatment of Osimertinib.

Keywords: *EGFR* recurrent mutation, Osimertinib, acquired resistance, *in silico* modeling, next generation sequencing

Introduction

Non-small cell lung cancer (NSCLC) represents one of the leading causes of cancer-related mortality worldwide [1], among which 10-50% patients bear activating mutations in epidermal growth factor receptor (*EGFR*), *e.g.* in-frame deletions in exon 19 (Ex19del) or missense mutation in exon 21 (L858R) [2]. The first (1st) and second (2nd) generations of EGFR tyrosine kinase inhibitors (TKIs) suffer from considerable drug resistance, *e.g. EGFR* secondary mutation T790M [3], targeting which the third (3rd) generation EGFR TKI Osimertinib (AZD9291) was approved by FDA in 2015 [4]. However, the majority of patients progressed on Osimertinib was found to develop newly acquired resistanct mutations, *e.g. EGFR* C797S, which destroys the covalent bonding with Osimertinib [5]. Recently, more and more *EGFR* mutations are discovered but their roles in Osimertinib resistance remain unknown [6].

Our early study of Osimertinib-resistant NSCLC patients revealed *EGFR* secondary mutations of L718 and L792, which are located in the p-loop and hinge of the ATP-binding site, respectively [7]. Further cellular experiments con-

firmed that both mutations were able to result in drug resistance against Osimertinib [8]. However, given the continual appearance of novel *EGFR* mutations and their extremely low frequency, using *in vitro* and/or *in vivo* experiments to predict drug sensitivity seems realistically forbidden. Recently, *in silico* methods, *e.g.* alchemical free energy calculation, have become promising in precision medicine for the patients carrying rare mutations [9]. For instance, Park *et al.* successfully applied a thermodynamic integration (TI) method to establish the role of the gatekeeper mutation on EGFR inhibitor selectivity [10]. Ikemura *et al.* reported that their established molecular dynamics (MD) simulation-based computational model, one of free energy perturbation (FEP) methods, predicted the diverse sensitivities of *EGFR* exon 20 insertion mutations to Osimertinib with surprisingly high consistency [11].

To explore the new resistant mechanisms, we analyzed the *EGFR* mutation profiles of 1,058 relapsed Chinese lung cancer patients after Osimertinib treatment, using targeted next-generation sequencing (NGS) with a comprehensive pan-cancer gene panel. Our study represents the largest *EGFR*-mutated lung cancer cohort so far investigating the Osimertinib resistance in a real-world setting. Furthermore, our *in silico* structural model was proved to be powerful and robust to predict the Osimertinib sensitivity of *EGFR* mutants.

Material and methods

Subjects

We retrospectively reviewed the genomic profiles of a cohort of 1058 lung cancer patients with acquired resistance to Osimertinib, a 3rd-generation EGFR tyrosine kinase inhibitors (TKIs) (Osimertinib-resistant group), and another cohort of 1803 lung cancer patients, who have only received 1st-generation EGFR-TKIs prior to disease relapse (1st-TKI-resistant group). Written informed consent was collected from each patient upon sample collection according to the protocols approved by the ethical committee of each hospital.

DNA extraction, hybridization capture, and sequencing

Blood samples were collected from the above patients after the disease progression upon drug treatment with Osimertinib or 1st generation TKIs. cfDNA was extracted using the NucleoSpin Plasma XS kit (Macherey Nagel). Purified DNA was qualified by Nanodrop2000 (Thermo Fisher Scientific) and quantified by Qubit 2.0 using the dsDNA HS Assay Kit (Life Technologies). Sequencing libraries were prepared using the KAPA Hyper Prep kit (KAPA Biosystems). A customized NGS panel sequencing was performed according to the previous report [12]. In brief, indexed DNA libraries were subjected to probe-based hybridization using IDT xGen Lockdown reagents (IDT, Coralville, IA) and Dynabeads M-270 (Thermo Fisher). Captured libraries were on-beads amplified with Illumina p5 and p7 primers in KAPA HiFiHotStartReadyMix (KAPA Biosystems). The final library was quantified by KAPA Library Quantification kit (KAPA Biosystems). Bioanalyzer 2100 (Agilent, Stanta Clara, CA) was used to determine the fragment size distribution of the final library. The target-enriched library was then sequenced on Illumina HiSeq4000 NGS platforms (Illumina) according to the manufacturer's instructions.

Sequence data processing and mutation calling

After demultiplexing, the sequencing data were subjected to Trimmomatic [13] for FASTQ file quality control (QC). Leading/trailing low quality (quality reading below 30) or N bases were removed. Qualified reads were mapped to the reference human genome hg19 using Burrows-Wheller Aligner (BWA-mem) [14]. Genome Analysis Toolkit (GATK) [15] was employed to apply the local realignment around indels and base quality score recalibration. PCR duplicates were removed by Picard. VarScan2 [16] was employed for the detection of single-nucleotide variations (SNVs) and insertion/deletion mutations with the recommended parameters [17]. We required minimum variant allele frequency of 0.3%, minimum variant supporting reads mapped to both strands of 3 for cfDNA samples. The resulted mutation lists were filtered through an internally collected list of recurrent sequencing errors on the same sequencing platform, which is summarized from the sequencing results of 400 normal blood samples. Specifically, if a variant was detected (i.e. ≥3 mutant reads) in >10% of the normal samples, it was considered a likely artifact and was

removed. As a result, 480 (Osimertinib-resistant group) and 929 patients (1st-TKI-resistant group) bearing *EGFR* mutants were included in the following analysis, individually.

Computational modeling and simulation

We used co-crystal structures of wild-type EGFR-KD in complex with Osimertinib (PDB ID: 4ZAU) and Gefitinib (PDB ID: 4WKQ) individually, to eliminate the potential changes in the protein structure upon drug binding (Osimertinib or Gefitinib) or other non-targeted mutation(s). The strategy by Ikemura *et al.* [9] was applied to build the *in silico* model of mutant EGFR-drug complex that each mutation was introduced into wild-type EGFR structure after modeling disordered loops and flexible side chains, which was also used in other studies of mutant kinase-drug interaction [18-21]. MD simulation and end-point binding free energy calculations were implemented following the protocols previously reported [22]. The computational details were delineated in the Supporting Information. (Computational Methods Section) To confirm the reliability of *in silico* modeling, we also collected experimentally determined IC_{50} (half inhibitory concentration) values for the specific mutants from our previous study [2], and compared the experimental results with our predictions.

EGFR mutation dynamics visualization

For *EGFR* mutation dynamics, we used PyClone [23] and SCHISM [24] to estimate cancer cell fractions and clonal architecture, which was visualized using the fishplot R package (https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-016-3195-z).

Statistical analysis

The Fisher's exact test was applied to *EGFR* alterations with >1% prevalence in the Osimertinib resistant group. P<0.05 was considered to indicate a statistically significant difference. All statistical analyses were performed with R packages.

Results

Overview of patients' characteristics

Here, we conducted a multicenter record review of two large cohorts of lung cancer patients who underwent genomic profiling at hospitals across China. The first cohort was comprised of a total of 1058 patients whose post-Osimertinib samples were included for analysis, regardless of prior TKI therapy other than Osimertinib. This cohort was delegated as the Osimertinib-resistant group and *EGFR* somatic mutations were detected in 480 patients. A total of 1803 patients who only received 1st-generation EGFR TKIs (*e.g.* Gefitinib, Erlotinib, Icotinib) prior to disease progression were allocated as the control group (1st-TKI-resistant group). Somatic *EGFR* mutations were detected in 929 patients of the 1st-TKI resistant group. A summary of the clinical characteristics of all patients was provided in Supplementary Materials (Table S1).

Identify EGFR mutations enriched in Osimertinib-resistant group

To identify *EGFR* mutations that may cause the resistance to Osimertinib rather than 1st-generation EGFR TKIs, we compared the mutational profiles of Osimertinib resistant patients with 1st-TKI-resistant group. For this, alterations were filtered by two criteria: 1) significantly enriched (P<0.05) in the Osimertinib-resistant group compared to the 1st-TKI-resistant group, and 2) rare but recurrent (N≥ 2) in Osimertinib-resistant group but not present in 1st-TKI-resistant group. We only focused on single amino acid substitution rather than complex mutations including large or small fragments insertion or deletion, for further MD simulation and structural analysis. The latter was believed to disturb the native protein structure more dramatically, therefore, it requires more extensive sampling and more accurate scoring for structural analysis.

Specifically, 7 mutations were found significantly enriched in the Osimertinib-resistant group compared with 1st-TKI-resistant group (**Figure 1**), including *EGFR* C797S mutation (20.6% vs. 0.4%, *P* value = 3.650818e-44), L718Q (4.6% vs. 0.2%, *P* value = 4.807101e-09), L792F (2.5% vs. 0.1%, *P* value = 2.229274e-06), L792H (1.9% vs. 0.2%, *P* value = 0.001574455), L718V (1.7% vs. 0.1%, *P* value = 0.0001745102), C797G (1.5% vs. 0.1%, *P* value = 0.0005172586), and G796S (1.3% vs. 0.1%, *P* value = 0.001531042), as shown in **Figure 1**. We also identified 13 recurrent mutations which were only observed in

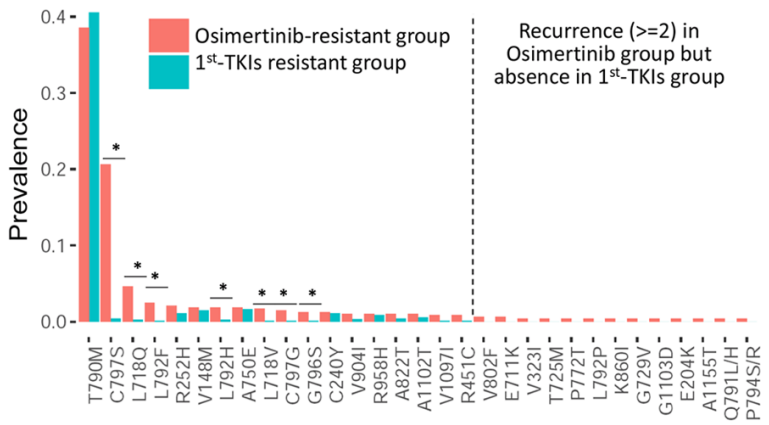


Figure 1. *EGFR* mutations enriched for Osimertinib resistance. *EGFR* mutations significantly enriched in the Osimertinib-resistant cohort (P<0.01) and 13 rare but recurrent (n≥2) mutations observed only in Osimertinib-resistant cohort.

Moreover, V802 is located at the first helix right next to the hinge region. Notably, 5 mutations (*e.g.* G796S, V802F, T725M, Q791L/H, P794S/R) were found co-existing with C797S (Figure S1). However, the retrospective inspection of sequence alignments showed that 10 patients, involving Q791L/H, P794S/R, G796S, V802F individually, all harbored the concurrent mutation *in trans* with *EGFR* C797S (Figure S2), suggesting their potential roles in drug resistance against Osimertinib.

Osimertinib-resistant group but not in 1st-TKI-resistant group (**Figure 1**), *e.g.* G796S (1.2% in prevalence, 6 patients), V802F (0.6% in prevalence, 3 patients), Q791L/H (0.4% in prevalence, 2 patients), P794S/R (0.4% in prevalence, 2 patients) (Table S2).

Energetic analysis of EGFR mutations on Osimertinib binding

To investigate the effects on Osimertinib binding, we applied a molecular dynamics simulation-based model to predict the sensitivity of rare *EGFR* mutations to 1st- and 3rd-generation EGFR-TKIs. To achieve a good balance between computational efficiency and accuracy, we focused on the mutations around the ATP-binding site in EGFR kinase domain (KD). For this, 7 mutational sites around the ATP-binding pocket, that is within 10 angstrom from Osimertinib in the co-crystal structure (PDB ID: 4ZAU), were selected for further analysis, including *EGFR* L718Q/V, L792F/H/P, T725M, Q791L/H, P794S/R, G796S, and V802F, among which the mutations at Q791 and P794, to the best of our knowledge, have not been reported previously.

The mutational sites that are located at the hinge region of EGFR KD, including *EGFR* Q791 (gatekeeper+1), L792 (gatekeeper+2), P794 (gatekeeper+4), G796 (gatekeeper+6), as shown in **Figure 2A**, were known to be essential for ATP binding [25]. In addition, both T725 and L718 are on the p-loop, which directly interact with small-molecule drugs (*e.g.* Osimertinib).

Our computational results showed that all of these EGFR mutations could reduce the binding energy of Osimertinib compared to that of wild-type (WT) EGFR evaluated by ΔΔG (**Figure 2B** and **Table 1**). The most dramatic changes were observed with Q791H, L718Q, T725M (range from 4.84 to 6.74 kcal/mol, **Table 1**), then followed by P794R/S, G796S, and Q791L (range from 2.76 to 4.21 kcal/mol, **Table 1**), while V802F and L792F were found with minimal effects on Osimertinib binding (range from 1.09 to 1.52 kcal/mol, **Table 1**). Similar results were observed in the 1st-generation TKI, Gefitinib, except that G796S was predicted to increase the drug binding affinity (ΔΔG: -2.37 kcal/mol, **Figure 2B** and **Table 1**).

Structural alterations underlying reduced drug binding affinity by EGFR mutations

To investigate the structural alterations underlying the reduced binding affinities by the above mutations, we compared the MD-simulated structural models with the native crystal structure and mainly focused on the drug binding as well as the protein context in the binding site. These EGFR mutations were predicted to interrupt the drug binding through various mechanisms including steric clash, hydrogen bond (HB) network alteration, and protein backbone strain elimination despite their close locations at the ATP-binding site.

Firstly, L718Q, L792F, and G796S all introduced steric clash which completely or partially impaired the key HB interaction with Osi-

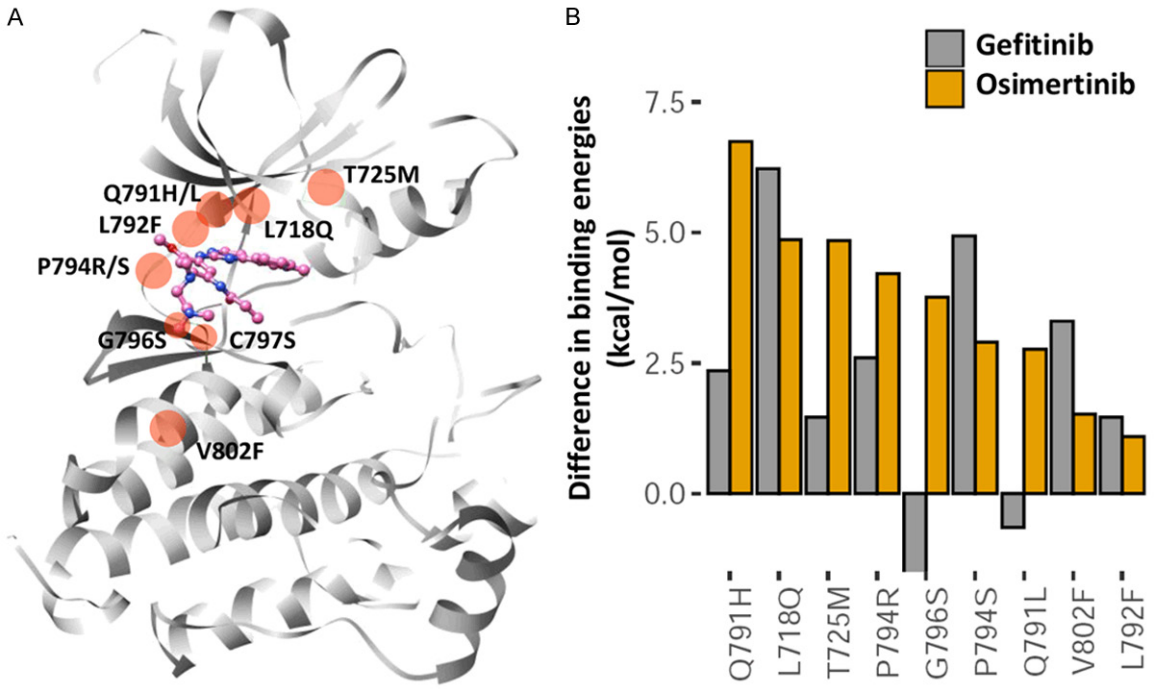


Figure 2. A. Structural view of identified Osimertinib-associated EGFR resistant mutations around ATP-site, where the drug bind. The kinase domain of EGFR protein is shown in grey ribbons with selected mutations highlighted with filled red circles. The small-molecule drug Osimertinib is represented with magenta ball-and-stick model. B. Difference in the calculated binding energies between mutant and wild-type EGFR proteins for each drug (yellow: Osimertinib, grey: gefitinib), a positive value implies unfavorable binding while negative value suggests favorable binding.

Table 1. Calculated energetic difference and root-mean-square deviation (RMSD) for small-molecule drugs in mutant and WT EGFR-KDs

Mutation	L718Q	L792F	G796S	Q791H/L	P794R/S	V802F	T725M
Osimertinib							
ΔΔG (kcal/mol) (vs. WT)	4.86	1.09	3.76	6.74/2.76	4.21/2.90	1.52	4.84
Ligand RMSD (Å) (vs. WT)	2.163	1.724	2.670	2.297/1.743	1.331/1.835	2.086	2.271
Gefitinib							
ΔΔG (kcal/mol) (vs. WT)	6.22	1.46	-2.37	2.35/-0.65	2.60/4.93	3.30	1.46
Ligand RMSD (Å) (vs. WT)	4.590	5.075	1.241	3.757/2.461	5.212/3.088	4.159	2.130

mertinib. Specifically, the bulkier and polar mutation (Gln) of L718Q is unfavorable to the binding of the aniline functional group in Osimertinib which results in steric hindrance thus destroying hydrophobic interaction with aromatic ring in comparison to the WT non-polar residue (Leu), giving rise to a significant deviation of the drug (ligand RMSD: 2.163 Å) (**Figure 3A**; **Table 1**). This is consistent with the loss of *in vitro* activity as previously described [26]. Similarly, L792F produced a slightly bulkier side-chain substitution (Phe) at the hinge region, which leads to the clashes with the selectivity-determined methoxyl functional group (ligand RMSD: 1.724 Å) although the key hydrogen bond (HB) is largely maintained (**Figure 3B**; **Table 1**). This also explained the residue sensitivity of L792F to Osimertinib in the previous cellular experiments [27]. G796 is also located at the hinge region (gatekeeper+6), whose backbone forms direct contacts with the aniline core. The substitution (Ser) introduces a hydroxyl in the side-chain, which is against the hydrophobic phenyl group and produces steric hindrance that pushes drug significantly away from the binding site (ligand RMSD: 2.670 Å) (**Figure 3C**; **Table 1**). Interestingly, gefitinib was shown to

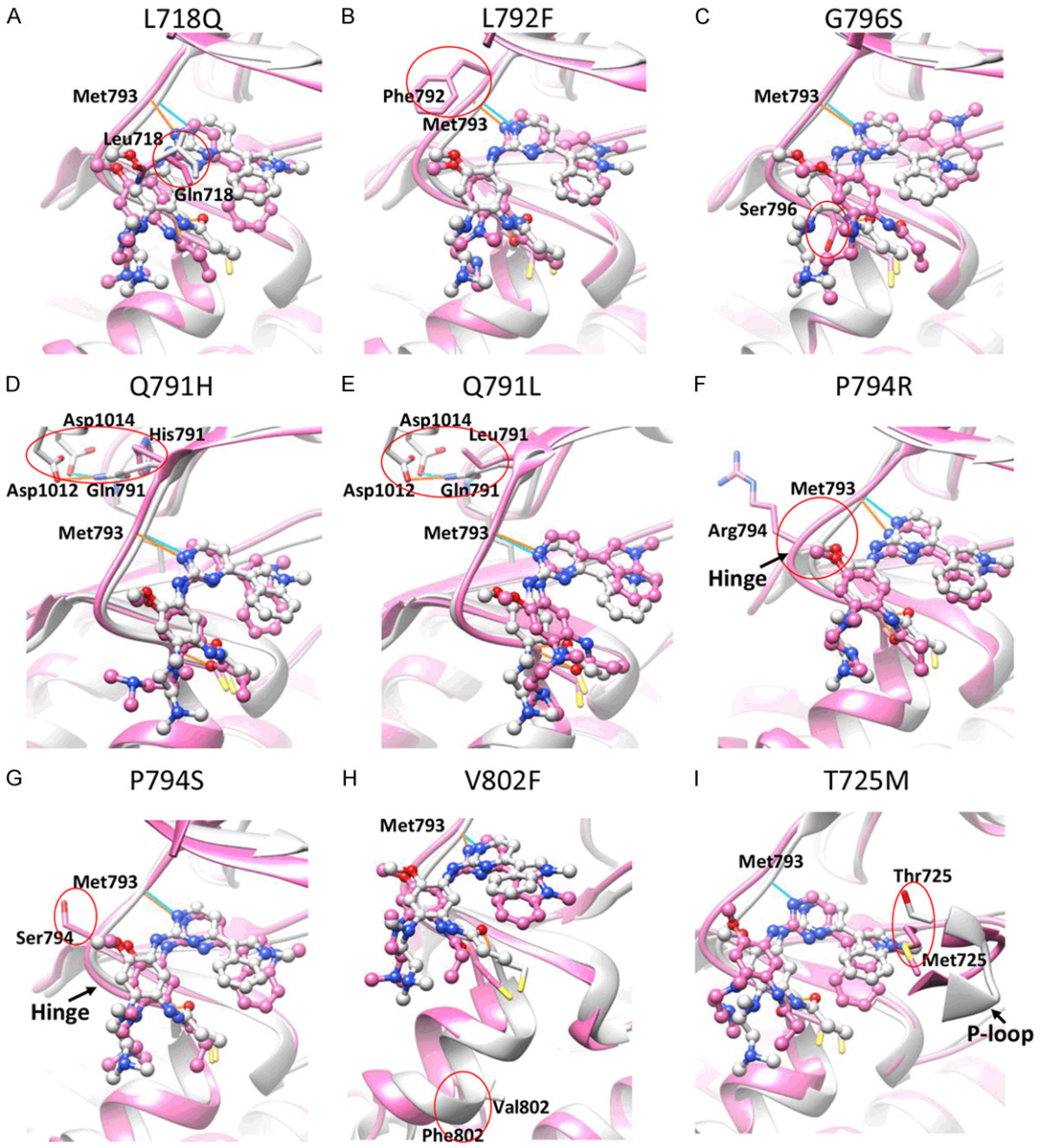


Figure 3. Structural insights into rare EGFR mutant variants bound with Osimertinib. Protein is shown in ribbon and stick for highlighted residues while drug molecule is represented with ball-and-stick model (white: crystal structure of WT EGFR, magenta: simulation structure of mutant EGFR). The selected mutations and induced structural changes are highlighted with red circles.

bind with EGFR G796S mutant even more favorably (ΔΔG: -2.37 kcal/mol, ligand RMSD: 1.241 Å) by forming additional HB with the hydroxyl group in mutated residue (Figure S3; **Table 1**).

Secondly, for the rare substitutions of Q791H/L, a conserved HB network was observed between the side-chains of Gln791 and Lys-852, Asp1012, Asp1014 in the WT EGFR, which however, was completely destroyed by these substitutions. This disturbed the binding of Osimertinib in the ATP-site, as suggested by the ligand RMSD values (Q791H: 2.297 Å, Q791L: 1.743 Å) (**Figure 3D** and **3E**).

The third predicted mechanism of reducing drug binding affinity was the elimination of

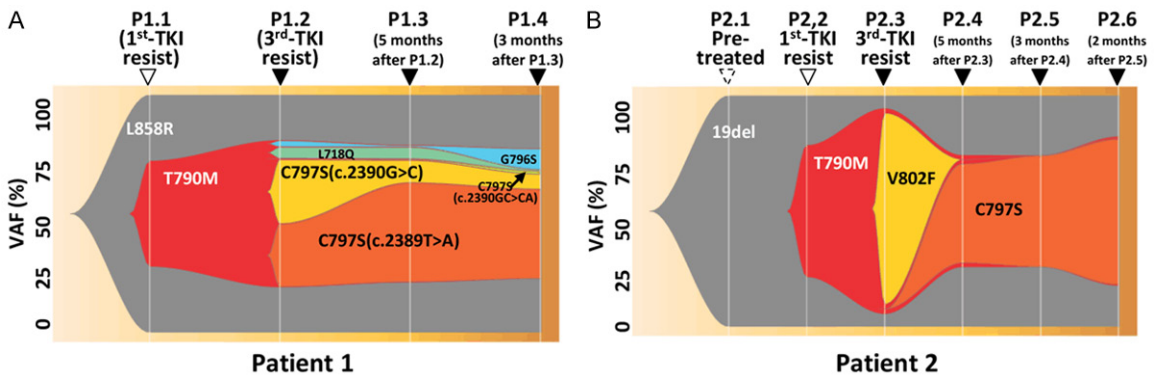


Figure 4. Fish plots of *EGFR* somatic variants detected in Osimertinib-resistant patients. Arrows indicate different time points for sampling, e.g. after acquiring drug resistance against the 1st-generation EGFR-TKI (empty) and 3rd-generation EGFR-TKI (filled), as well as the pre-treatment (empty and dash line). Y-axis represents VAF (%) and X-axis shows the time progression.

native backbone strain by P794R/S which was introduced by the conserved hinge region [28]. The proline-produced backbone strain is believed to be important for the maintenance of the protein secondary and tertiary structure [29], therefore the mutations could be expected to destabilize the drug binding (ligand RMSD: 1.331 Å for P794R, 1.835 Å for P794S) (**Figure 3F** and **3G**).

Other structural changes were observed in the rare mutations V802 and T725 though they were located relatively far from the drug binding site. The simulation results showed that V802F mutation displaced the first helix right next to the hinge region compared to that in WT EGFR, as suggested by the atomic distance of backbone Ca (residue RMSD: 2.590 Å). This movement disturbed drug binding (ligand RMSD: 2.086 Å) (**Figure 3H**). Similarly, the bulkier and polar-to-hydrophobic mutation of T725M displaced the p-loop by moving towards the ATP-site, which resulted in the conflict with the drug (ligand RMSD: 2.271 Å) (**Figure 3I**).

Despite all rare mutations are located around the ATP-binding site, the large diversity of structural alterations underlying Osimertinib resistance present a great challenge to structure-based drug design of the next-generation EGFR-TKIs targeting ATP-binding site, which is known with a few common features, *e.g.* interaction with hinge region (with 1 or 2 hydrogen bonds), direct contacts with p-loop, and others. Encouragingly, as suggested by the increased binding affinity of Geftinib and G796S, a proper but unexpected drug repurposing can be achieved. This is of practical use in a clinical setting, which however, demands elaborate identification of cross-activity of different TKIs across a wide variety of resistance-associated mutant EGFRs.

V802F and G796S are acquired EGFR rare mutations coexistent with C797S

Considering the contribution of *EGFR* rare mutations to Osimertinib resistance, we were interested in exploring the mutation dynamics related to the drug resistance clinical course. With the retrospective inspection of our cohort, especially those with available electronic health record data, 2 patients who received Osimertinib treatment after acquiring 1st-TKI resistance, were characterized by the multiple samples collected at different time points, including pre- and post-treatment of Osimertinib. This allowed us to track EGFR genetic alterations within 2 years with four analyzed samples in the first patient (P1) and for over 2 years in the second patient (P2) with 6 samples.

As shown in **Figure 4A**, the coexistence of *EGFR* rare mutation G796S with C797S was observed in Patient 1 (P1). Specifically, P1 was initially treated with the 1st-generation EGFR-TKI (Icotinib) for 1.5 years and then received the 3rd-generation EGFR-TKI (Osimertinib) when the progressed disease (PD) occurred. The classic EGFR oncogenic driver mutation L858R and 1st-generation-TKI-resistance mutation T790M were detected in the

first sample (P1.1) collected upon Icotinib resistance. However, the sequential NGS analysis revealed multiple sub-clonal mutations (C797S, G796S, and L718Q) acquired after Osimertinib treatment. In addition, 3 different DNA-level variations were detected all leading to *EGFR* C797S mutation, including c.2389T>A (VAF: 22.4%), c.2390G>C (VAF: 22.8%), and c.2390GC>CA (VAF: 0.1%). In October 2018 (P1.3), mutational data indicated an increase of c.2389T>A (VAF: 37.6%) and a decrease of c.2390G>C (VAF: 8.3%) for *EGFR* C797S. The same mutations were also detected in P1.4 except for a considerable increase in the *EGFR* G796S (VAF: 5.7%), suggesting the contribution of G796S to Osimertinib resistance as well as the coexistence of dominant (C797S) and rare (L718Q, G796S) mutations. Of note, the 1st-generation EGFR-TKI, Gefitinib, was predicted to maintain the binding activity towards G796S mutant EGFR, suggesting the susceptibility to Gefitinib among patients with acquired resistance to first-line Osimertinib.

As shown in **Figure 4B**, the sequential occurrence of C797S after *EGFR* rare mutation V802F was observed in patient 2 (P2). A total of 6 samples from P2 were obtained for NGS analysis including treatment-naïve, post-Icotinib, and post-Osimertinib. In sample P2.2 (December 2016, pre-Osimertinib sample), *EGFR* T790M was detected with a VAF of 29.2%. With disease progression, another sample was analyzed in August 2017 (P2.3) and we were able to detect two new subclonal *EGFR* mutations, V802F and C797S with VAFs of 77.5% and 0.1%, respectively. The higher VAF of V802F suggested its potential role in drug resistance. In January 2018 (P2.4), mutational data indicated a dramatic increase in the *EGFR* C797S (VAF: 41.0%), accompanied by the significant decrease of *EGFR* V802F. Different from P1, a single DNA-level alteration, c.2389T>A, was observed to cause *EGFR* C797S in P2. *EGFR* C797S lasted for the following time points (P2.5 and P2.6). In a word, our findings provided an additional view of mutation dynamics during the time course of EGFR-TKIs treatment.

Discussion

In this study, we provided both genomic and structural insights into the novel mechanism of Osimertinib-resistant EGFR mutations using a very large cohort of lung cancer patients. Several rare mutations associated with drug resistance were identified, including T725M, Q791L/H, and P794S/R. All of them were characterized with a low frequency of <0.5%. Considering the recent approval of Osimertinib as the first-line treatment of metastatic NSCLC [29], the expanded use would inevitably result in more unrecognized drug-resistant mutations. This can be evidenced by the recent retrospective analysis of the FLAURA trial, which revealed that *EGFR* C797S was only present in 7% of patients acquiring resistance in the studied cohort [30]. Moreover, the landscape of drug resistance for Osimertinib is largely unknown. Our work has disclosed a collection of *EGFR* rare mutations enriched in Osimertinib resistant patients, which could serve as guidance in the clinical practice. Our findings also suggested *EGFR* C797S, as a dominant mutation, is likely to coexist with other rare mutation(s) (*e.g. EGFR* G796S and V802F) among a small proportion of patients progressed on Osimertinib treatment, which together, contribute to the drug resistance. The coexistence of EGFR rare mutation(s) with C797S observed in the present study suggested the mechanism of resistance to Osimertinib might be heterogeneous.

Considering the extremely low frequency of *EGFR* rare mutation and the continual appearance of novel mutations, experimental evaluation of drug sensitivity is realistically forbidden in a clinical setting. Encouragingly, *in silico* methods hold promise in providing alternative solutions. Several studies turned to the alchemical free energy calculation, *e.g.* FEP and TI, and proved the power in precision medicine for the patients carrying rare mutations [9-11]. However, the extremely high computational cost largely prevents practical use, which demands the assistance of supercomputers or clusters. In contrast, the endpoint method is shown to achieve a good balance between efficiency and accuracy [26]. In the present work, we predicted both *EGFR* L718Q and L792F mutants are able to reduce the binding affinities of Osimertinib and Gefitinib, and L718Q renders more resistance to both drugs than L792F. These results are in agreement with our previous experimental findings that both *EGFR* mutants (L718Q and L792F) render

increased IC_{50} values for two drugs compared with *EGFR* without the specific mutation, as well as L718Q presents much higher IC_{50} values than L792F for Osimertinib (L718Q: >1000 nM, L792F: 72.41 nM) and Gefitinib (L718Q: 500 nM, L792F: 31.25 nM), respectively [2]. Our results suggested the endpoint free energy calculation, *e.g.* MM/PBSA, is able to delineate drug sensitivity across a large diversity of EGFR mutations by simply estimating and comparing the binding free energies. Herein, we only focused on the single point mutation rather than more complex ones, *e.g.* small/large fragment insertion/deletion, which is believed to disturb the native protein structure more dramatically, and therefore, demands much more expensive computations.

The atomic-level MD simulations could provide structural insights into the mechanism of drug resistance. Despite the very close locations within or around ATP-binding site, many different roles were still seen for the studied rare mutations, *e.g.* introducing steric clash and impairing key HB interaction with hinge region (L718Q, L792F, G796S), disturbing protein native HB network (Q791H/L), eliminating imposed backbone strain in the native structure (P794R/S), and other effects. These together, present a great challenge for structure-based drug design, because most of small-molecule TKIs are known to target ATP-binding site, the large diversity of structural mechanisms observed for rare mutations in the present study suggest an unrealistic optimism in developing EGFR-TKI with ability to overcome all these resistant mutation effects simultaneously. Notably, our computational study found that compared with EGFR WT, Gefitinib even prefers to bind with EGFR G796S which shows resistance against Osimertinib, indicating the potential therapeutic strategy for this drug-resistant mutation. However, a proper but unexpected drug repurposing can only be achieved by elaborating the comprehensive profiles of cross-activity of different TKIs across a wide variety of resistant mutations, where *in silico* methods would be of great practical use.

Acknowledgements

This work was supported by grants from the Key Research and Development plan (Social development) of science and technology department of Jiangsu Province (No. BE2019-760), the Medical Innovation Team Foundation of the Jiangsu Provincial Enhancement Health Project (No. CXTDA2017021).

Disclosure of conflict of interest

Ran Cao, Qiuxiang Ou, Yutong Ma, Hua Bao, and Xue Wu are employees of Geneseeq Technology Inc., Toronto, Canada. Yang Shao is an employee of Nanjing Geneseeq Technology Inc., Nanjing, Jiangsu, China.

Address correspondence to: Dr. Bo Shen, Department of Oncology, Jiangsu Cancer Hospital, Jiangsu Institute of Cancer Research, Nanjing Medical University Affiliated Cancer Hospital, Nanjing, Jiangsu, China. Tel: +86-13913910555; E-mail: Shenbo987@126.com; Dr. Zhaoxia Wang, Department of Oncology, The Second Affiliated Hospital of Nanjing Medical University, Nanjing, Jiangsu, China. Tel: +86-18951762628; E-mail: zhaoxiawang66@126.com

References

[1] Global Burden of Disease Cancer Collaboration, Fitzmaurice C, Dicker D, Pain A, Hamavid H, Moradi-Lakeh M, MacIntyre MF, Allen C, Hansen G, Woodbrook R, Wolfe C, Hamadeh RR, Moore A, Werdecker A, Gessner BD, Te Ao B, McMahon B, Karimkhani C, Yu C, Cooke GS, Schwebel DC, Carpenter DO, Pereira DM, Nash D, Kazi DS, De Leo D, Plass D, Ukwaja KN, Thurston GD, Yun Jin K, Simard EP, Mills E, Park EK, Catalá-López F, deVeber G, Gotay C, Khan G, Hosgood HD 3rd, Santos IS, Leasher JL, Singh J, Leigh J, Jonas JB, Sanabria J, Beardsley J, Jacobsen KH, Takahashi K, Franklin RC, Ronfani L, Montico M, Naldi L, Tonelli M, Geleijnse J, Petzold M, Shrime MG, Younis M, Yonemoto N, Breitborde N, Yip P, Pourmalek F, Lotufo PA, Esteghamati A, Hankey GJ, Ali R, Lunevicius R, Malekzadeh R, Dellavalle R, Weintraub R, Lucas R, Hay R, Rojas-Rueda D, Westerman R, Sepanlou SG, Nolte S, Patten S, Weichenthal S, Abera SF, Fereshtehnejad SM, Shiue I, Driscoll T, Vasankari T, Alsharif U, Rahimi-Movaghar V, Vlassov VV, Marcenes WS, Mekonnen W, Melaku YA, Yano Y, Artaman A, Campos I, MacLachlan J, Mueller U, Kim D, Trillini M, Eshrati B, Williams HC, Shibuya K, Dandona R, Murthy K, Cowie B, Amare AT, Antonio CA, Castañeda-Orjuela C, van Gool CH, Violante F, Oh IH, Deribe K, Soreide K, Knibbs L, Kereselidze M, Green M, Cardenas R, Roy N,

Tillmann T, Li Y, Krueger H, Monasta L, Dey S, Sheikhbahaei S, Hafezi-Nejad N, Kumar GA, Sreeramareddy CT, Dandona L, Wang H, Vollset SE, Mokdad A, Salomon JA, Lozano R, Vos T, Forouzanfar M, Lopez A, Murray C and Naghavi M. The Global Burden of Cancer 2013. JAMA Oncol 2015; 1: 505-527.

[2] Yang Z, Yang N, Ou Q, Xiang Y, Jiang T, Wu X, Bao H, Tong X, Wang X, Shao YW, Liu Y, Wang Y and Zhou C. Investigating novel resistance mechanisms to third-generation EGFR tyrosine kinase inhibitor osimertinib in non-small cell lung cancer patients. Clin Cancer Res 2018; 24: 3097-3107.

[3] Kobayashi S, Boggon TJ, Dayaram T, Janne PA, Kocher O, Meyerson M, Johnson BE, Eck MJ, Tenen DG and Halmos B. EGFR mutation and resistance of non-small-cell lung cancer to gefitinib. N Engl J Med 2005; 352: 786-792.

[4] Pao W, Miller VA, Politi KA, Riely GJ, Somwar R, Zakowski MF, Kris MG and Varmus H. Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain. PLoS Med 2005; 2: e73.

[5] Greig SL. Osimertinib: first global approval. Drugs 2016; 76: 263-273.

[6] Thress KS, Paweletz CP, Felip E, Cho BC, Stetson D, Dougherty B, Lai Z, Markovets A, Vivancos A, Kuang Y, Ercan D, Matthews SE, Cantarini M, Barrett JC, Janne PA and Oxnard GR. Acquired EGFR C797S mutation mediates resistance to AZD9291 in non-small cell lung cancer harboring EGFR T790M. Nat Med 2015; 21: 560-562.

[7] Niederst MJ, Hu H, Mulvey HE, Lockerman EL, Garcia AR, Piotrowska Z, Sequist LV and Engelman JA. The allelic context of the C797S mutation acquired upon treatment with third-generation EGFR inhibitors impacts sensitivity to subsequent treatment strategies. Clin Cancer Res 2015; 21: 3924-3933.

[8] Minari R, Bordi P and Tiseo M. Third-generation epidermal growth factor receptor-tyrosine kinase inhibitors in T790M-positive non-small cell lung cancer: review on emerged mechanisms of resistance. Transl Lung Cancer Res 2016; 5: 695-708.

[9] Ikemura S, Yasuda H, Matsumoto S, Kamada M, Hamamoto J, Masuzawa K, Kobayashi K, Manabe T, Arai D, Nakachi I, Kawada I, Ishioka K, Nakamura M, Namkoong H, Naoki K, Ono F, Araki M, Kanada R, Ma B, Hayashi Y, Mimaki S, Yoh K, Kobayashi SS, Kohno T, Okuno Y, Goto K, Tsuchihara K and Soejima K. Molecular dynamics simulation-guided drug sensitivity prediction for lung cancer with rare EGFR mutations. Proc Natl Acad Sci U S A 2019; 116: 10025-10030.

[10] Okada K, Araki M, Sakashita T, Ma B, Kanada R, Yanagitani N, Horiike A, Koike S, Oh-Hara T, Watanabe K, Tamai K, Maemondo M, Nishio M, Ishikawa T, Okuno Y, Fujita N and Katayama R. Prediction of ALK mutations mediating ALK-TKIs resistance and drug re-purposing to overcome the resistance. EBioMedicine 2019; 41: 105-119.

[11] Park J, McDonald JJ, Petter RC and Houk KN. Molecular dynamics analysis of binding of kinase inhibitors to WT EGFR and the T790M mutant. J Chem Theory Comput 2016; 12: 2066-2078.

[12] Ruan Z, Katiyar S and Kannan N. Computational and experimental characterization of patient derived mutations reveal an unusual mode of regulatory spine assembly and drug sensitivity in EGFR kinase. Biochemistry 2017; 56: 22-32.

[13] Bolger AM, Lohse M and Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014; 30: 2114-2120.

[14] Lu H, Yang S, Zhu H, Tong X, Xie F, Qin J, Han N, Wu X, Fan Y, Shao YW and Mao W. Targeted next generation sequencing identified clinically actionable mutations in patients with esophageal sarcomatoid carcinoma. BMC Cancer 2018; 18: 251.

[15] Li H and Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. Bioinformatics 2009; 25: 1754-1760.

[16] McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M and DePristo MA. The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010; 20: 1297-1303.

[17] Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L and Wilson RK. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res 2012; 22: 568-576.

[18] Hauser K, Negron C, Albanese SK, Ray S, Steinbrecher T, Abel R, Chodera JD and Wang L. Predicting resistance of clinical Abl mutations to targeted kinase inhibitors using alchemical free-energy calculations. Commun Biol 2018; 1: 70.

[19] Aldeghi M, Gapsys V and de Groot BL. Predicting kinase inhibitor resistance: physics-based and data-driven approaches. ACS Central Science 2019; 5: 1468-1474.

[20] Aldeghi M, Gapsys V and de Groot BL. Accurate estimation of ligand binding affinity changes upon protein mutation. ACS Central Sci 2018; 4: 1708-1718.

[21] Liu J, Pei J and Lai L. A combined computational and experimental strategy identifies mutations conferring resistance to drugs targeting the BCR-ABL fusion protein. Commun Biol 2020; 3: 18.
[22] Cao R, Liu M, Yin M, Liu Q, Wang Y and Huang N. Discovery of novel tubulin inhibitors via structure-based hierarchical virtual screening. J Chem Inf Model 2012; 52: 2730-2740.
[23] Roth A, Khattra J, Yap D, Wan A, Laks E, Biele J, Ha G, Aparicio S, Bouchard-Cote A and Shah SP. PyClone: statistical inference of clonal population structure in cancer. Nat Methods 2014; 11: 396-398.
[24] Niknafs N, Beleva-Guthrie V, Naiman DQ and Karchin R. SubClonal hierarchy inference from somatic mutations: automatic reconstruction of cancer evolutionary trees from multi-region next generation sequencing. PLoS Comput Biol 2015; 11: e1004416.
[25] Wang Y, Sun Y, Cao R, Liu D, Xie Y, Li L, Qi X and Huang N. In silico identification of a novel hinge-binding scaffold for kinase inhibitor discovery. J Med Chem 2017; 60: 8552-8564.
[26] Cao R, Huang N and Wang Y. Evaluation and application of MD-PB/SA in structure-based hierarchical virtual screening. J Chem Inf Model 2014; 54: 1987-1996.
[27] Rousseau F, Schymkowitz JW, Wilkinson HR and Itzhaki LS. Three-dimensional domain swapping in p13suc1 occurs in the unfolded state and is controlled by conserved proline residues. Proc Natl Acad Sci U S A 2001; 98: 5596-5601.
[28] McHarg J, Kelly SM, Price NC, Cooper A and Littlechild JA. Site-directed mutagenesis of proline 204 in the 'hinge' region of yeast phosphoglycerate kinase. Eur J Biochem 1999; 259: 939-945.
[29] Scott LJ. Osimertinib as first-line therapy in advanced NSCLC: a profile of its use. Drugs Ther Perspect 2018; 34: 351-357.
[30] Soria JC, Ohe Y, Vansteenkiste J, Reungwetwattana T, Chewaskulyong B, Lee KH, Dechaphunkul A, Imamura F, Nogami N, Kurata T, Okamoto I, Zhou C, Cho BC, Cheng Y, Cho EK, Voon PJ, Planchard D, Su WC, Gray JE, Lee SM, Hodge R, Marotti M, Rukazenkov Y and Ramalingam SS; FLAURA Investigators. Osimertinib in untreated egfr-mutated advanced non-small-cell lung cancer. N Engl J Med 2018; 378: 113-125.

Computational methods

In-silico modeling of EGFR-TKI complex

The crystal structure of human EGFR kinase domain (KD) in complex with osimertinib, the third-generation tyrosine kinase inhibitor (3^{rd}-TKI), was obtained from the Protein Data Bank (PDB ID: 4ZAU). The disordered loops and flexible side chains were modeled using Modeller v9.2 program *(1)*. As a comparison, we also used the crystal structure of human EGFR-KD in complex with gefitinib (1^{st}-TKI, PDB ID: 4WKQ) as our computational control. For the selected mutations, each was introduced into the crystal structure of wild-type (WT) EGFR with fully structural optimizations to eliminate any potential induced steric clashes.

System setup and MD simulation

MD simulations were performed with AMBER12.0 suite *(2)*. AMBER99SB force field was applied for the protein, and the general Amber force field (GAFF) was applied for the small molecules. ANTECHAMBER was used for calculating the AM1-BCC charges of small molecules. All system setups were performed in TLEAP and simulations were carried out using SANDER, during which 3 stages of minimization were performed in the gas phase, and then followed by the addition of a 30 Å water cap based on the geometric center of binding site. After 200 ps of equilibration of solvents, a production run of 5 ns was performed on the whole system at 300 K with a time step of 2.0 fs. All residues including solvents beyond 12 Å of small molecules are fully frozen.

Endpoint binding energy calculation

Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PB/SA) calculation was carried out following the protocol previously reported *(3)*. Generally, 100 snapshots were evenly extracted from last 1 ns MD simulation. The interaction energy between protein and small molecules as well as the internal energy were calculated with the SANDER. As for the solvation energy, the polar contribution is calculated using PBSA with PARSE radii, while the nonpolar part is estimated proportional to the solvent-accessible surface area. The difference in the binding energies can be estimated with the formula: $\Delta\Delta G = \Delta G_{Mutant} - \Delta G_{WildType} = (G_{complex} - G_{receptor} - G_{ligand})_{Mutant} - (G_{complex} - G_{receptor} - G_{ligand})_{WildType}$.

References

[1] Sali A and Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 1993; 234: 779-815.

[2] Case DA, Cheatham TE 3rd, Darden T, Gohlke H, Luo R, Merz KM Jr, Onufriev A, Simmerling C, Wang B and Woods RJ. The amber biomolecular simulation programs. J Comput Chem 2005; 26: 1668-88.

[3] Cao R, Liu M, Yin M, Liu Q, Wang Y and Huang N. Discovery of novel tubulin inhibitors via structure-based hierarchical virtual screening. J Chem Inf Model 2012; 52: 2730-40.

Table S1. Clinical characteristics of patients with identified EGFR somatic mutations

Characteristics	Group I (n = 480) Osimertinib_resist	Group II (n = 929) 1st-TKI-only_resist	*p*-value
Male (%)	201 (41.9%)	380 (40.9%)	0.7322
Female (%)	279 (58.1%)	549 (59.1%)	
Age at diagnosis (Median)	24-87 (61) years	21-96 (62) years	>0.01

Table S2. Clinical characteristics of patients with identified EGFR somatic mutations

Mutations	Group I (n = 480) Osimertinib_resist		Group II (n = 929) 1st-TKI-only_resist		*p*-value
	Prevalence	No. of patients	Prevalence	No. of patients	
T790M	0.385	185	0.406	377	
C797S	0.206	99	0.004	4	3.65E-44
L718Q	0.046	22	0.002	2	4.81E-09
L792F	0.025	12	0.001	1	2.23E-06
R252H	0.021	10	0.011	10	
V148M	0.019	9	0.015	14	
L792H	0.019	9	0.002	2	1.57E-03
A750E	0.019	9	0.016	15	
L718V	0.017	8	0.001	1	0.000175
C797G	0.015	7	0.001	1	0.000517
G796S	0.013	6	0.001	1	0.001531
C240Y	0.013	6	0.011	10	
V904I	0.010	5	0.003	3	
R958H	0.010	5	0.009	8	
A822T	0.010	5	0.004	4	
A1102T	0.010	5	0.005	5	
V1097I	0.008	4	0.001	1	
R451C	0.008	4	0.001	1	
V802F	0.006	3	0.000	0	
E711K	0.006	3	0.000	0	
V323I	0.004	2	0.000	0	
T725M	0.004	2	0.000	0	
P772T	0.004	2	0.000	0	
L792P	0.004	2	0.000	0	
K860I	0.004	2	0.000	0	
G729V	0.004	2	0.000	0	
G1103D	0.004	2	0.000	0	
E204K	0.004	2	0.000	0	
A1155T	0.004	2	0.000	0	
Q791L/H	0.004	2	0.000	0	
P794S/R	0.004	2	0.000	0	

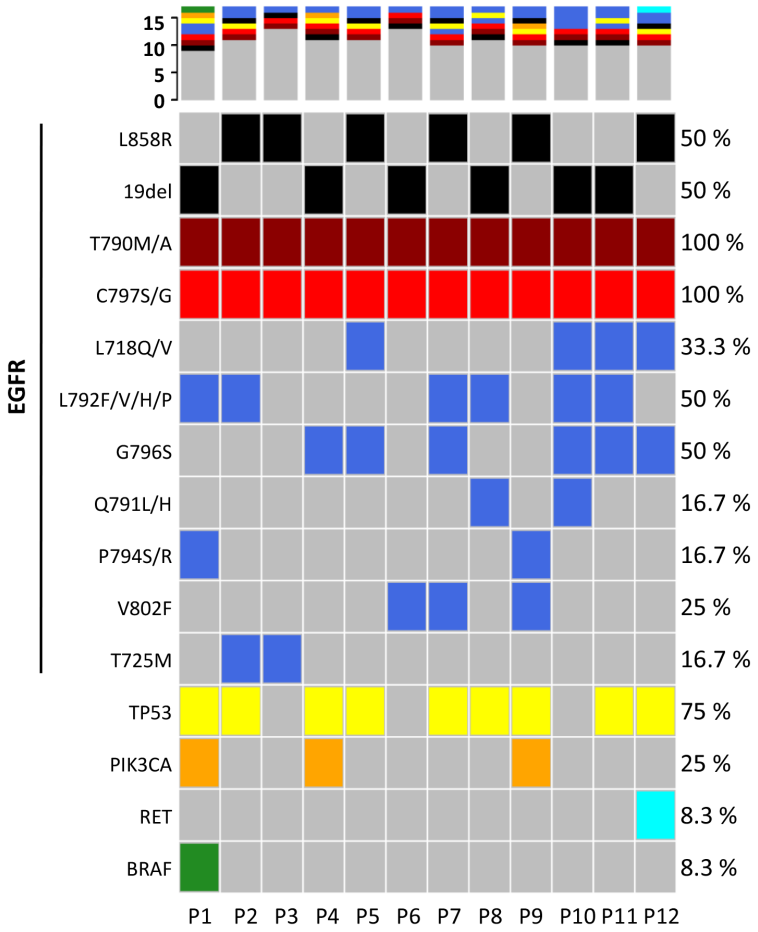
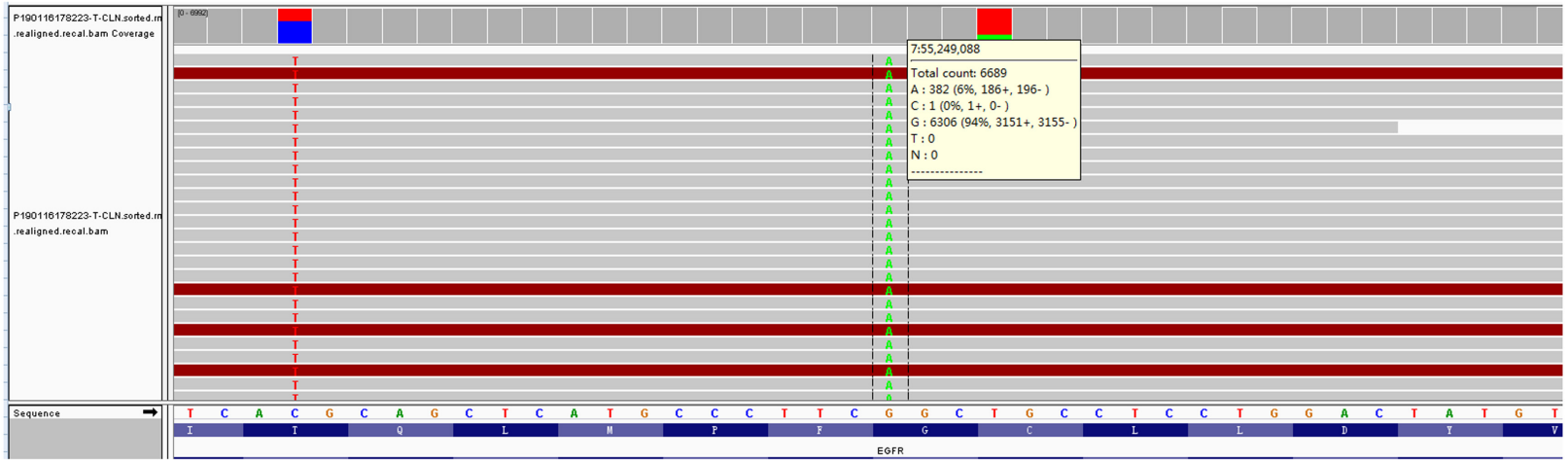
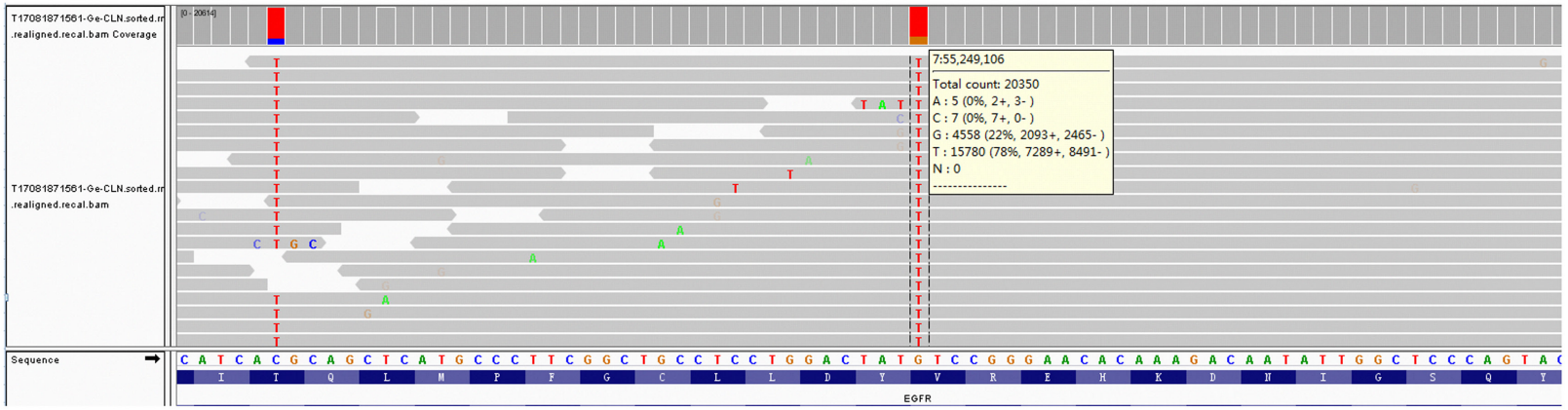


Figure S1. Highlighted EGFR and other driver genetic mutations detected among 12 selected lung cancer patients. The original oncogenic mutations (L858R and 18del) are colored in black, the major resistant EGFR mutations at T790 and C797 are colored in rosybrown (1st-TKI resistant) and red (3rd-TKI resistant), the minor resistant EGFR mutations at L718, L792, G796, Q791, P794, V802, T725 are colored in skyblue, while mutations in other selected genes are colored in yellow (TP53), orange (PIK3CA), cyan (RET), forest green (BRAF).

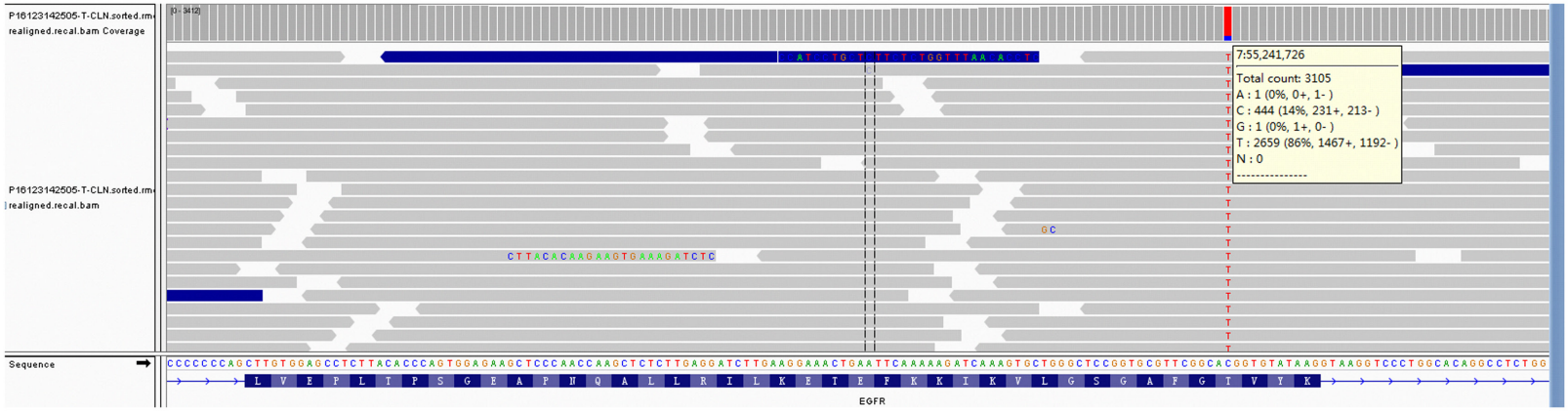
G796S


P190116178223-T-CLN.sorted.rn
.realigned.recal.bam Coverage
P190116178223-T-CLN.sorted.rn
.realigned.recal.bam
7:55,249,088
Total count: 6689
A : 382 (6%, 186+, 196-)
C : 1 (0%, 1+, 0-)
G : 6306 (94%, 3151+, 3155-)
T : 0
N : 0
Sequence
EGFR

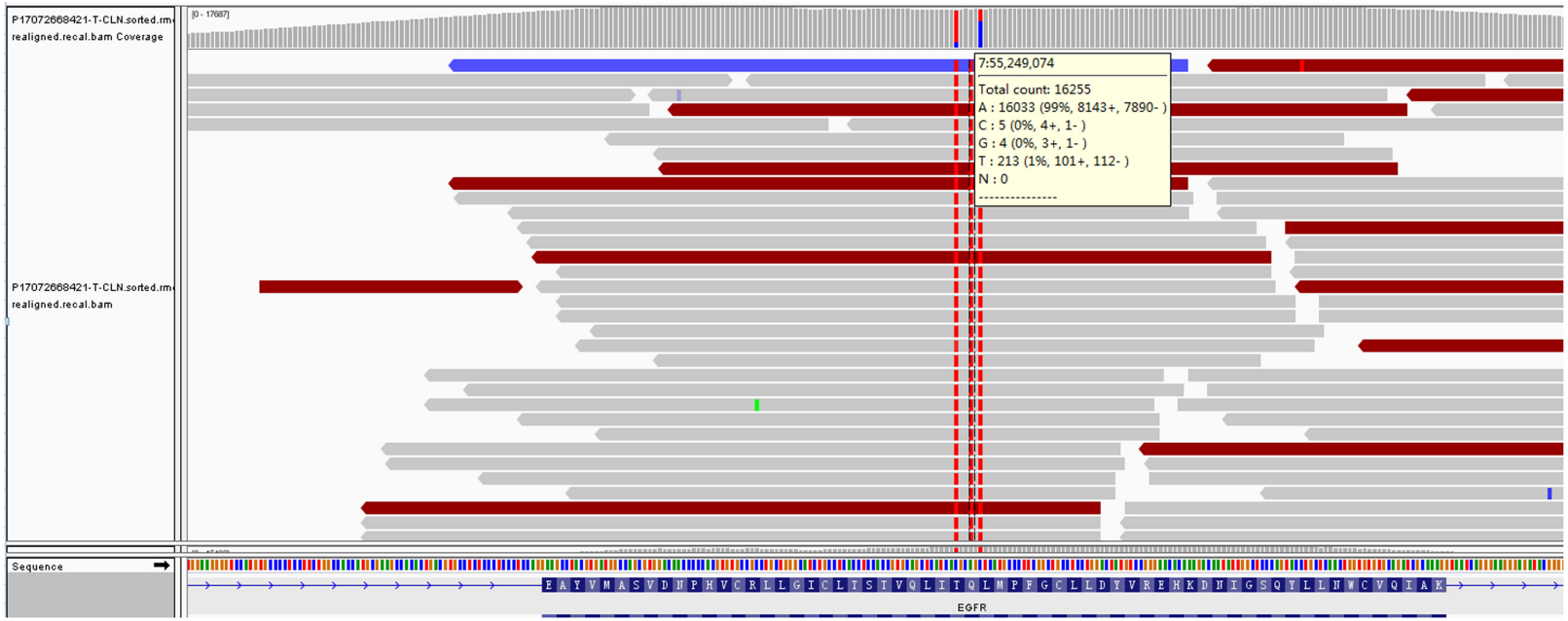
V802F


T17081871561-Ge-CLN.sorted.rn
.realigned.recal.bam Coverage
T17081871561-Ge-CLN.sorted.rn
.realigned.recal.bam
7:55,249,106
Total count: 20350
A : 5 (0%, 2+, 3-)
C : 7 (0%, 7+, 0-)
G : 4558 (22%, 2093+, 2465-)
T : 15780 (78%, 7289+, 8491-)
N : 0
Sequence
EGFR

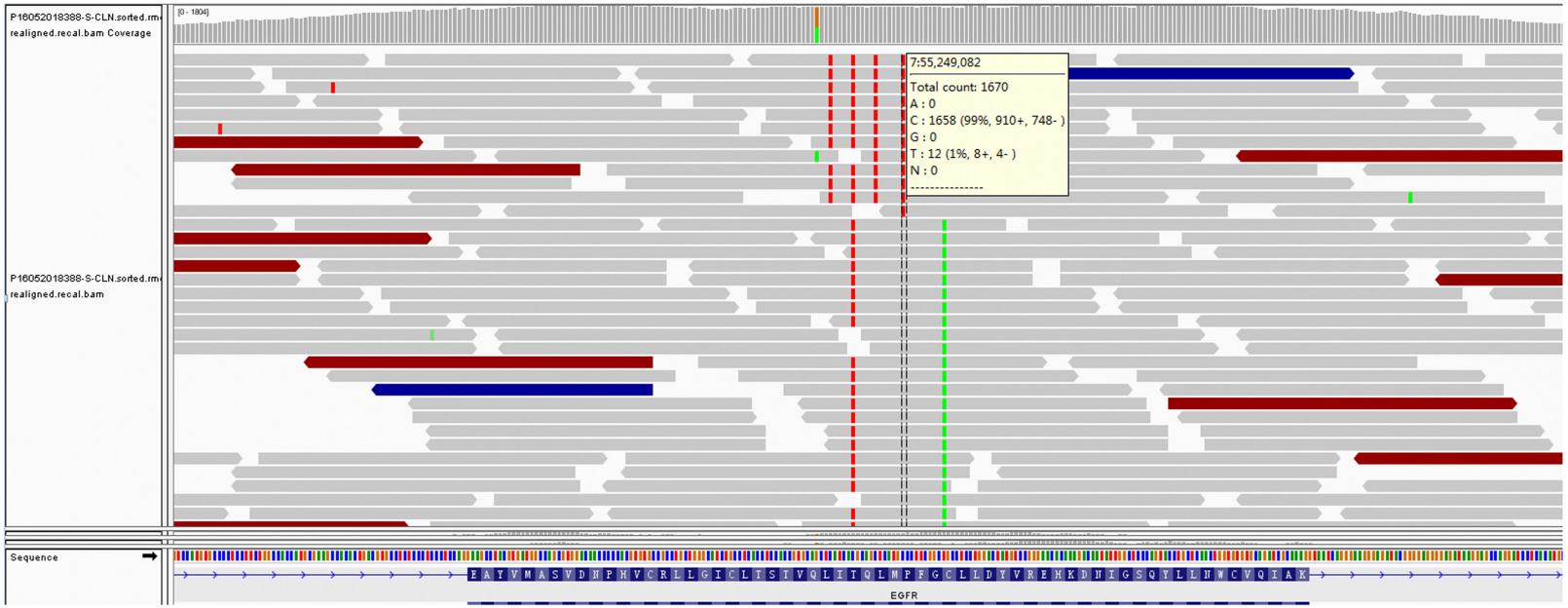
T725M


P16123142505-T-CLN.sorted.rm
realigned.recal.bam Coverage
P16123142505-T-CLN.sorted.rm
realigned.recal.bam
7:55,241,726
Total count: 3105
A : 1 (0%, 0+, 1-)
C : 444 (14%, 231+, 213-)
G : 1 (0%, 1+, 0-)
T : 2659 (86%, 1467+, 1192-)
N : 0
Sequence
EGFR

Q791L


P17072668421-T-CLN.sorted.rm
realigned.recal.bam Coverage
P17072668421-T-CLN.sorted.rm
realigned.recal.bam
7:55,249,074
Total count: 16255
A : 16033 (99%, 8143+, 7890-)
C : 5 (0%, 4+, 1-)
G : 4 (0%, 3+, 1-)
T : 213 (1%, 101+, 112-)
N : 0
Sequence
EGFR

P794S


P16052018388-S-CLN.sorted.rm
realigned.recal.bam Coverage
[0 - 1804]
7:55,249,082
Total count: 1670
A : 0
C : 1658 (99%, 910+, 748-)
G : 0
T : 12 (1%, 8+, 4-)
N : 0
P16052018388-S-CLN.sorted.rm
realigned.recal.bam
Sequence
EGFR

P794R

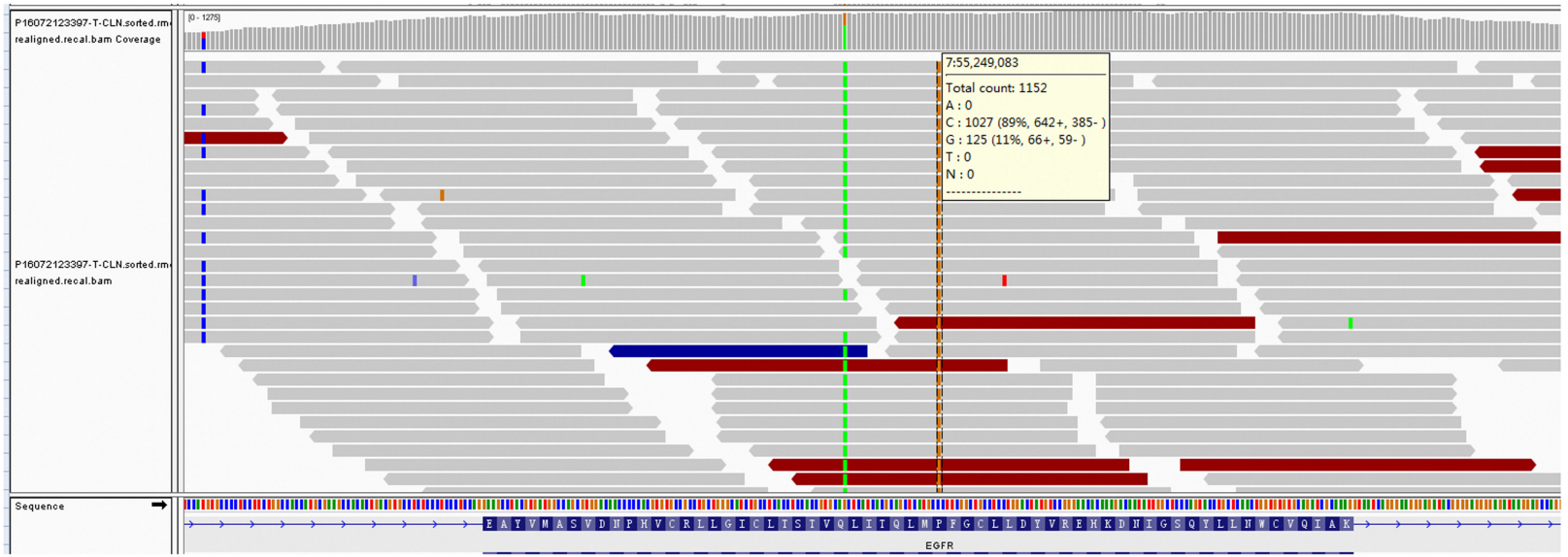

P16072123397-T-CLN.sorted.rm
realigned.recal.bam Coverage
[0 - 1275]
7:55,249,083
Total count: 1152
A : 0
C : 1027 (89%, 642+, 385-)
G : 125 (11%, 66+, 59-)
T : 0
N : 0
P16072123397-T-CLN.sorted.rm
realigned.recal.bam
Sequence
EGFR

Figure S2. Sequence alignment for the selected patients with a focus on the EGFR resistant mutations around ATP-binding site.

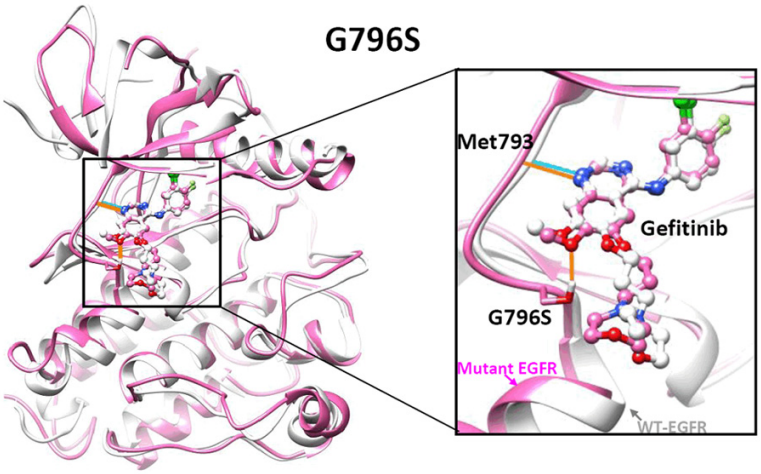


Figure S3. Structural insights into the effects of EGFR-G796S mutation on gefitinib binding by comparing the complex structure of WT and mutant EGFR bound with drug. Protein is shown in ribbon and stick for highlighted residues while drug molecule is represented with ball-and-stick model (white: crystal structure of WT EGFR, magenta: simulation structure of mutant EGFR). The hydrogen bonds were shown in cyan and orange lines.