Article

Deciphering the Impact of Rare Missense Variants in EGFR-TKI-Resistant Non-Small-Cell Lung Cancer through Whole Exome Sequencing: A Computational Approach

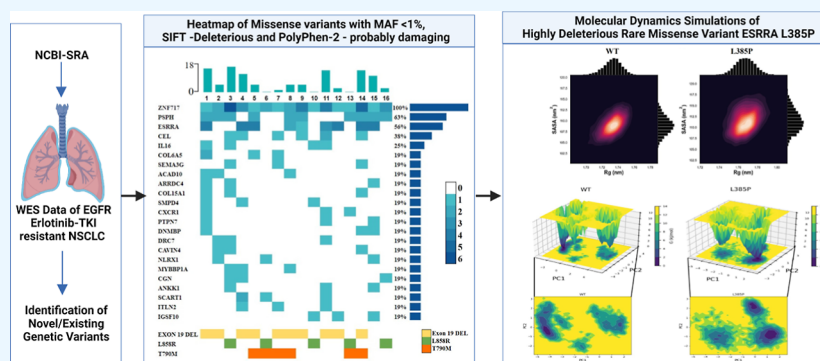Ambritha Balasundaram and George Priya C Doss*

Read Online

ACCESS | Metrics & More | Article Recommendations | SI Supporting Information

ABSTRACT: Targeted therapy revolutionizes the treatment of non-small-cell lung cancer (NSCLC), harboring molecular change. Epidermal growth factor receptor(*EGFR*) mutations play a crucial role in the development of NSCLC, serving as a pivotal factor in its pathogenesis. We elucidated the mechanisms of resistance and potential therapeutic strategies in NSCLC resistant to the EGFR-tyrosine kinase inhibitor (EGFR-TKI). This is achieved by identifying rare missense variants through whole exome sequencing (WES). The goal is to enhance our understanding, identify biomarkers, and lay the groundwork for targeted interventions, thereby offering hope for an improved NSCLC treatment landscape. We conducted WES analysis on 16 NSCLC samples with EGFR-TKI-resistant NSCLC obtained from SRA-NCBI (PRJEB50602) to reveal genomic profiles within the EGFR-TKI. Our findings showed that 48% of the variants were missense, and after filtering with the Ensembl variant effect predictor, 53 rare missense variants in 23 genes were identified as highly deleterious. Further examination using pathogenic tools like PredictSNP revealed 12 deleterious rare missense variants in 7 genes: *ZNF717, PSPH, ESRRA, SEMA3G, PTPN7, CAVIN4,* and *MYBBP1A.* Molecular dynamics simulation (MDS) suggested that the L385P variant alters the structural flexibility of ESRRA, potentially leading to unfolding of ERRα proteins. This could impact their function and alter ERRα expression. These insights from MDS enhance our understanding of the structural and dynamic consequences of the L385P ESRRA variant and provide valuable implications for subsequent therapeutic considerations and targeted interventions.

1. INTRODUCTION

Lung cancer ranks as the second most frequently diagnosed cancer and the primary cause of cancer-related deaths globally. According to the Global Cancer Observatory (GLOBOCAN), about 2.2 million (11.4%) people were newly diagnosed with lung cancer in 2020, with 1.8 million (18%) deaths reported.[1] As per the GLOBOCAN data for India in 2020, approximately 72,510 individuals, constituting 5.5% of the total, prevailing new lung cancer diagnoses. The confirmed deaths associated with lung cancer amounted to 66,279, representing 7.8% of all cancer-related deaths. The five-year prevalence across all age groups reached 80,817. Lung cancers are typically categorized into two primary types: small-cell lung cancer and non-small-cell lung cancer (NSCLC). NSCLC constitutes approximately 80−85% of all lung cancer cases and can be further classified

into lung adenocarcinoma, squamous cell carcinoma (SCC), and large cell carcinoma. Patients with metastatic cancer have a 5 year survival rate of only 6%.[2]
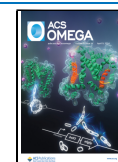
The presence of oncogene drivers in NSCLC indicates the existence of a distinct subgroup of patients who stand to gain substantial benefits from targeted therapy, particularly in advanced and metastatic stages. Individuals with actionable gene mutations undergoing treatment with targeted agents

exhibited a noteworthy enhancement in overall survival rates and a reduced mortality risk compared to counterparts who did not receive targeted treatment or lacked oncogene addiction.[3] With the investigation into activating mutations within the tyrosine kinase domain of epidermal growth receptor (EGFR), there has been a burgeoning identification of activating oncogenic alterations in NSCLC. These alterations include *ALK*, *ROS-1*, *BRAF*, *RET*, *MET*, and *ERBB2*, each with obvious implications for treatment choices and patient survival. Notably, some of these alterations have received approvals for NSCLC treatments from regulatory bodies such as the Food and Drug Administration and the European Medicines Agency.[4] Despite the remarkable efficacy of these targeted medications, all patients eventually experience disease progression due to inherent or acquired resistance mechanisms.[5]

The mutated *EGFR* typically indicates the tumor's susceptibility to tyrosine kinase inhibitors (TKIs). Regrettably, tumor recurrence or the emergence of TKI resistance is a common occurrence. Consequently, three generations of TKIs have been developed in response to this challenge.[6] Gefitinib and erlotinib, the initial EFGR TKIs of the first generation, have demonstrated remarkable efficacy in NSCLC patients with *EGFR* exon 19 deletions/L858R mutations. Nevertheless, primary TKI resistance is observed in approximately 10% of NSCLC cases. Furthermore, nearly half of the patients treated with EGFR-TKIs develop T790M mutations upon resistance to TKI treatment.[7,8] Second-generation EGFR-TKIs, including afatinib, neratinib, and dacomitinib, demonstrate inhibitory action against *EGFR* by forming covalent bonds with the Cys797 residue, leading to irreversible kinase inhibition. These TKIs have demonstrated increased anticancer efficacy compared to first-generation TKIs. However, their effectiveness is limited due to the frequent development of drug resistance, particularly when targeting the T790M *EGFR* mutation.[9] Osimertinib, an irreversible third-generation oral EGFR-TKI, exhibits potent and selective inhibition of *EGFR* with exon 19 deletions/L858R mutations and *EGFR* T790M resistance mutations. Clinical trials have demonstrated the effectiveness of Osimertinib in individuals with NSCLC possessing *EGFR* with exon 19 deletions/L858R and T790M mutations, even those with central nervous system metastases. It has demonstrated enhanced progression-free survival (PFS) compared to other EGFR-TKIs like erlotinib or gefitinib.[10] The emergence of acquired *EGFR* C797S, an on-target resistance mechanism observed after exposure to irreversible inhibitors, is frequently noted in 24% of patients receiving a second-line or subsequent Osimertinib treatment. Acquired mutation rates for *PIK3CA*, *KRAS*, and *BRAF* were typically in line with the rates reported in the second-line Osimertinib treatment.[10−12] In addition to *EGFR* T790M, the reported acquired resistance mechanisms, including *MET* amplification and *HER2* amplification, align with the resistance mechanisms observed in response to first- and second-generation EGFR-TKIs. Diverse resistance mechanisms emerged in patients treated with Osimertinib.[13,14] Understanding individual tumor heterogeneity and the evolutionary pressures linked to different therapies have proven highly valuable in shaping future treatment strategies for patients with oncogene-addicted NSCLC.[15] The rapid, high-throughput, and precise nucleic acid insights provided by next-generation sequencing have significantly advanced our understanding of cancer biology and expanded therapeutic possibilities.[16] Whole exome sequencing

(WES) focuses on the protein-coding region of the genome, allowing the identification of variants impacting protein. This targeted approach, capturing key disease-causing mutations, reduces sequencing costs, making exome sequencing a clinically viable option for patient diagnostics.[17]

In light of the significant obstacle posed by drug resistance in NSCLC, there is an urgent imperative to explore the disparities of the EFGR-TKI resistance mechanisms and characterize the genomic patterns of EGFR-TKI-resistant patients using WES. Additionally, the aim is to identify the deleterious rare missense variants in EGFR-TKI-resistant patients through in silico pathogenicity assessment. Subsequently, molecular dynamics simulations (MDSs) will study the structural and functional impact of the identified deleterious variant.

2. MATERIALS AND METHODS

2.1. Data Collection and Preprocessing. We searched the NCBI-SRA database for WES data sets of *EGFR*-mutated TKI resistant NSCLC samples.[18] From PRJEB50602, we found the 16 pair-end FASTQ reads of WES data with *EGFR*-mutated erlotinib TKI-resistant lung adenocarcinoma tissue samples (*Homo sapiens*), which were processed from the Illumina HiSeq 2500 platform.[19] These raw reads were further downloaded from the database using a sequence read archive (SRA) toolkit.[20] Supplementary Figure 1 depicts the entire study workflow. The FASTQ files comprise raw reads originating from millions of DNA fragments. The data were refined employing filters to eliminate adaptor sequences, unwanted contaminant sequences, and low-quality bases. The trimmomatic tool accurately clips contaminant sequences and low-base quality bases from the reads.[21] The fundamental quality of the raw sequence data was then checked using the FastQC tool (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/).

2.2. Variant Calling, Filtering, and Annotation. The sequence reads were aligned to the reference human genome GRCH38 using the Burrows-Wheeler Aligner−Maximal Exact Matches (BWA−MEM).[22] This alignment resulted in a sequence alignment/map file compressed to its binary counterpart, a binary alignment/map (BAM) file using the Samtools view. Subsequently, it sorted the BAM file using Samtools sort. Samtools mpileup was then used to generate the mpileup file from the sorted BAM file. Employing a BCFtools call on the pileup file generated an output file in the variant call format (VCF) to call the variants. The SAMtools variant caller was employed in this study to identify single nucleotide polymorphisms (SNPs) and insertions/deletions.[23] Unlike various aligners and variant callers, BWA-MEM and SAMtools exhibited superior performance, particularly in Illumina data sets. The mutations obtained were documented in VCF or binary call format files, serving as inputs for subsequent analyses, such as the annotation of variant effects on encoded proteins. Functional annotations were performed against the Ensembl database using the Ensembl variant effect predictor (VEP).[24] The VEP tool provided information on genes and transcripts affected by the variants, their locations, coding consequences, and additional insights from tools like SIFT, PolyPhen-2, and gnomAD allele frequency (AF). The SIFT tool predicts the impact of an amino acid substitution on protein function. The prediction is based on the physicochemical similarity of the alternative amino acids and sequence homology. Each substitute receives a grade and a qualitative assessment as either "tolerated (≥0.05)" or "deleterious

(<0.05)". The PolyPhen-2 tool estimates the impact of an amino acid change on the structure and function of a protein using sequence homology, Pfam annotations, 3D structures, and several other databases. For each amino acid substitution, PolyPhen-2 provides a qualitative prediction and a score such as probably damaging (>0.908), possibly damaging (>0.446 and ≤0.908), benign (≤0.446), or unknown.

2.3. In Silico Evaluation of Rare Missense Variants. The PredictSNP web server was used to characterize the functional nonsynonymous single nucleotide polymorphisms (nsSNPs). Compiled from three distinct data sets by rectifying discrepancies, duplications, and previously used mutations, it utilized a standard data set of over 43,000 mutations for unbiased evaluation. Eight widely recognized prediction tools were assessed: nsSNPAnalyzer, MAPP, PANTHER, PolyPhen-1, PhD-SNP, PolyPhen-2, SIFT, and SNAP. The six top-performing tools were integrated into a consensus classifier, PredictSNP. This approach significantly improved prediction accuracy and provided rapid results for all mutations, demonstrating that consensus prediction offers a reliable and robust alternative to individual tool predictions.[25]

2.4. Biophysical Characteristic, Evolutionary Conservation, and Protein Stability Analysis. A missense substitution's biological effects can be correctly predicted using Align-GVGD, a biophysical characterization method that integrates protein multiple sequence alignments with the biophysical characteristics of amino acids. A GVGD assesses trans-activity by evaluating Grantham variation (GV), which quantifies biochemical diversity within amino acids, and Grantham deviation (GD), indicating the biochemical gap of the mutant amino acid from the observed one at a specific position. The tool categorizes substitutions based on their GD values, predicting their GD values as neutral, deleterious, or unclassified. A GD of 0 (GV > 61.3 and 0 < GD ≤ 61.3) is considered neutral. Conversely, the missense substitution is deemed deleterious if GV = 0 and GD is greater than 0.[26] Seven classes were established from them. C0 is probably neutral or benign, and C15, C25, C35, C45, CSS, and C65 are probably deleterious or pathogenic.[27] This analysis provides insights into the potential impact of missense substitutions on the protein function.

The ConSurf online tool assessed the evolutionary conservation of amino acid positions by assigning a rating between 1 (indicating the most changeable position) and 9 (indicating the most conserved position) for each position.[28]

The assessment of protein stability was conducted through I-Mutant2.0, a tool that anticipates the influence of amino acid substitutions on protein stability, providing an $\Delta\Delta G$ value (free energy change). The $\Delta\Delta G$ value represents the energy difference between the folded and unfolded protein structures. A positive $\Delta\Delta G$ value suggests that the mutation enhances protein stability, while a negative value indicates decreased stability.[29] This analysis aids in identifying the potential influence of amino acid changes on the overall stability of the protein.

2.5. Prediction of Harmful Variants. We utilized the MutPred tool to analyze the target protein sequence in the FASTA format and predict the deleterious effects of rare missense variants. The MutPred tool, designed for predicting the impact of single-site mutations as disease-related or neutral in humans, relies on protein structure, function, and evolution. MutPred2 scores range between 0 and 1, with a higher score indicating a greater propensity to be pathogenic. It employs a

random forest method, providing a "g" score for deleterious substitutions and a "p" score for molecular mechanism distribution.[30]

2.6. Molecular Dynamics and Simulation. The MDS employed the CHARMM27 force field within the GROMACS package version 2021.2.[31,32] The crystal structure of the estrogen-related receptor α (ESRRA or ERRα) protein (PDB ID: 3K6P) and the wild-type (WT) protein structure was used, with the water and ligand coordinates removed and filled missing residues.[33] Variant L385P protein structures were generated from the WT structure using the mutate tool option in the Swiss-Pdb viewer. The systems were solvated with a cube-shaped container of TIP3P water molecules, and neutralization was achieved by adding 48 Na$^+$ and 39 Cl$^-$ counterions with water molecules. The steepest descent energy minimization was applied to each system following neutralization until the maximum force dropped below 500 kJ mol^{-1} nm^{-1}. Two position-restrained MDS (*NVT* and *NPT*) were then conducted to equilibrate the ions and solvent surrounding the proteins. The temperature and pressure of the systems were controlled at 300 K and 1 bar, respectively, by using a Parrinello–Rahman barostat and V-rescale thermostat. Subsequently, each system underwent an unrestrained MDS (MD production run) lasting 500 ns with the same parameters as those of the previous simulations with positional constraints. Hydrogen atom bonds were constrained with the LINCS algorithm, and long-range electrostatic interactions were managed using the particle mesh Ewald method.[34,35] The cutoff distances for Lennard-Jones and short-range and long-range electrostatic interactions were uniformly set to 12 Å. Integration of Newton's equations of motion was performed with a time step of 2 fs.

2.7. Trajectory Analysis and Visualization. Trajectory analyses were conducted using the GROMACS package 2021.2. Several parameters, including root-mean-square deviation (rmsd), root-mean-square fluctuation (RMSF), the radius of gyration (R_g), solvent-accessible surface area (SASA), and intramolecular hydrogen bonds (H-bonds), were evaluated using specific GROMACS tools such as gmx rmsd, gmx rmsf, gmx gyrate, gmx sasa, and gmx hbond. Principal component analysis (PCA) was performed using gmx covar and gmx anaeig. For free energy landscape (FEL) analysis, the all-atom backbone of the protein was calculated. It utilized the gmx sham of the GROMACS to compute the Gibbs free energy and construct the FEL. The secondary structural element (SSE) pattern was estimated using the (gmx dssp) DSSP program within the GROMACS.

3. RESULTS

3.1. Exome Sequencing Screening the Rare Missense Variants. The downloaded paired-end WES FASTA files of 16 NSCLC with EGFR-TKI erlotinib-resistant samples were analyzed to identify rare missense variants. The Supplementary Table 1 contains the details of the sample. We meticulously evaluated the sequenced reads, both before and after the trimming of low-quality reads using FastQC. Subsequently, we aligned the reads to the GRCH38 reference genome using a BWA-MEM aligner, achieving a consistent overall alignment rate of 85% in each sample. We then executed indexing, sorting, and variant calling operations using Samtools. The resulting VCF file was then enhanced with annotations using the VEP. The WES analysis found an average of 243464.4375 single nucleotide variations in the 16 NSCLC samples. We
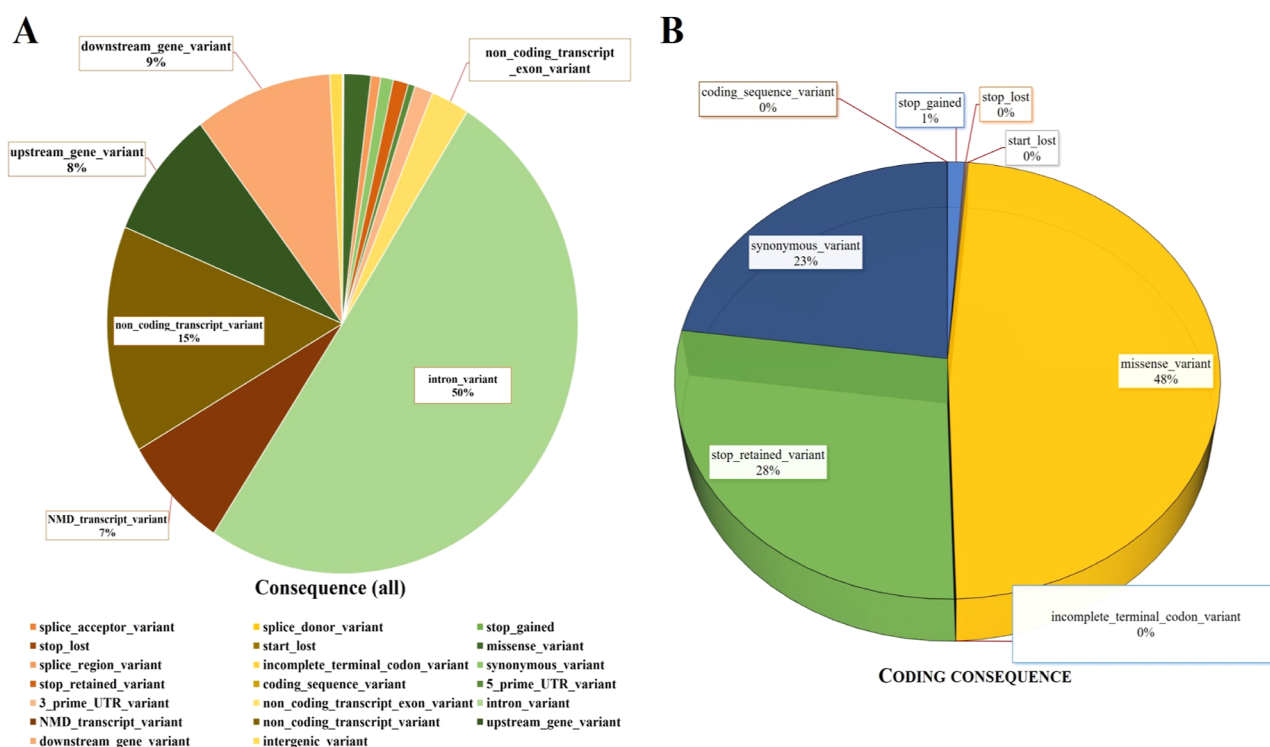
Figure 1. (A) Pie chart represents the average of 16 NSCLC samples. Consequence variant types that are involved in the overall NSCLC EGFR resistance samples. (B) Pie chart represents the average of 16 NSCLC samples. Coding consequence variant distributions in percentage.

have identified an average of 101974.5 (41.8%) novel and 167066.9 (68.6%) existing genetic variants in 16 NSCLC with EGFR-TKI erlotinib-resistant samples. Supplementary Table 2 contains each sample's total variants, including novel and existing variants in numbers and also mentioned in percentages. The distribution of consequence variants in 16 NSCLC EGFR-TKI-resistant samples, on average, is presented in Figure 1A. Similarly, the overall coding consequence variants obtained from this study were averaged and shown in Figure 1B. Among these coding consequence variants, we screened the missense variants with gnomAD AF less than 1%, SIFT as deleterious (<0.05), and PolyPhen as probably damaging (>0.908) in each sample to identify the deleterious missense rare variants of *EGFR*-mutated TKI-resistant NSCLC patients. We further prioritized and constructed the heatmap from the obtained deleterious missense variant-encoded genes by considering if the genes were present in more than or equal to three samples, which were significantly missense rare variant-enriched genes in these data. After prioritization, we found 53 deleterious rare missense variants in 23 genes, as illustrated in Figure 2. Detailed information on these genes and their variants, along with the gnomAD AF values, SIFT, and PolyPhen-2 scores, are provided in supplementary Table 3.

3.2. Prioritizing the Functional Rare Missense Variants. We used the PredictSNP tool, which utilizes the nsSNPAnalyzer, MAPP, PolyPhen-1, PhD-SNP, PolyPhen-2, SIFT, SNAP, and PANTHER, to determine the deleterious variants that can significantly alter the structure or function of the proteins. Out of 53 rare variants, 12 variants in 7 genes were deemed deleterious in 7 or 8 tools (Table 1). The nsSNPAnalyzer tool showed unknown results for all the rare variants. Align GVGD predicts the functional impact on protein as deleterious in seven rare variants belonging to C65 (Table 2). Consurf predicted that six rare variants are located

in evolutionarily most conserved positions (score = 9) of the proteins (Table 2). I-Mutant2.0 predicted that 11 rare variants showed a decreased stability of the proteins upon substitutions, whereas 1 rare variant (*ESRRA* L385P) showed increased stability (Table 2). From our analysis, three rare variants were found to be highly deleterious, which were further validated using MutPred2. It predicts the pathogenicity of amino acid substitutions and their molecular mechanisms. MutPred2 predicted that *ESRRA* "L385P" obtained a higher score (0.961), which indicates a greater propensity to be pathogenic than the others (Table 3).

3.3. MDS Trajectory Analysis. *3.3.1. rmsd and RMSF Analyses.* To investigate the impact of L385P mutation on the molecular functions of ESRRA protein, MDS lasting 500 ns were carried out for the ESRRA wild-type (WT) and L385P. rmsd of the backbone atoms was generated to assess the convergence of the MD trajectories. The rmsd plot revealed that the WT and L385P ESRRA proteins reached stable trajectories after 150 ns (Figure 3A). The rmsd mean and standard deviation of the WT and L385P were 0.33 ± 0.04 and 0.28 ± 0.02 nm, respectively (Figure 3B). The graph indicates that the L385P mutation has a lower rmsd than the WT, suggesting a minimal difference in the average values. rmsd analysis leads us to conclude that our simulation results are consistent and reliable for further dynamic property analysis.

RMSF analysis was performed to examine the broad overview of the protein's flexible regions and to assess the impact of L385P mutation on protein residue dynamics. We observed reduced RMSF values around the R1 (285−300 residues) and R3 (400−410 residues) regions of L385P than the WT, indicating that mutation reduced these regions' flexibility. Subsequently, we found higher RMSF values around the R2 (365−390 residue) region of L385P than in the WT, suggesting that mutation increased the flexibility in this region
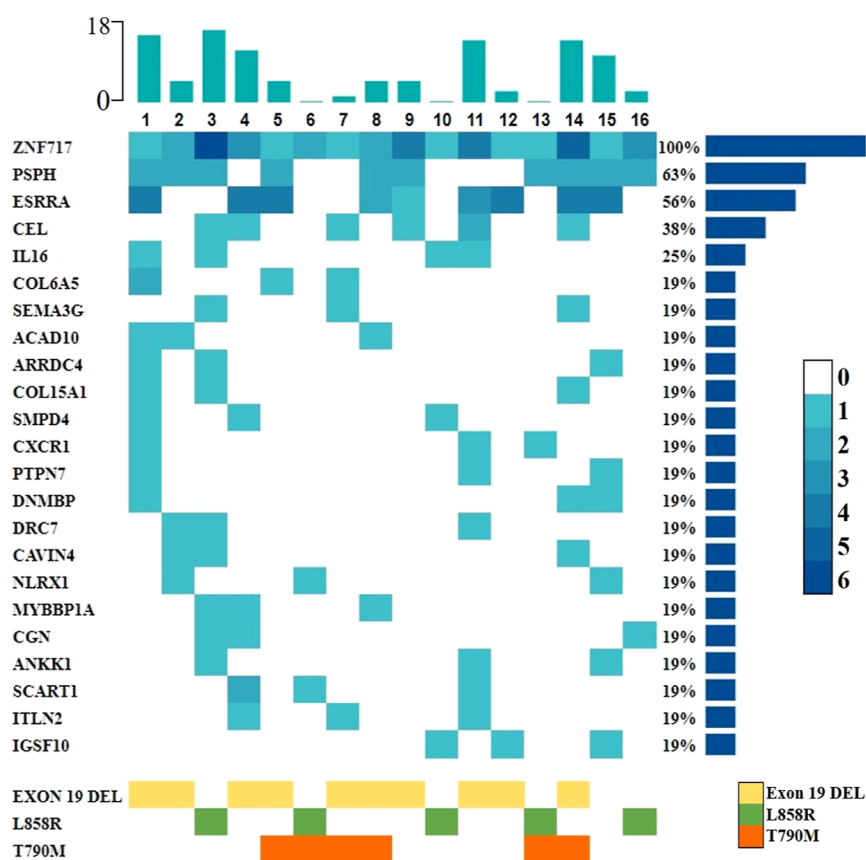
Figure 2. Heatmap of 53 deleterious rare missense variants across 23 genes in the WES of 16 EGFR-TKI-resistant NSCLC patients. Each row corresponds to a gene, and each column corresponds to the sample ID of NSCLC patients. Here, the gradient from darker blue to light blue signifies the number of deleterious rare missense variants within a specific gene. Horizontal bar graphs display the percentage of genes containing deleterious rare variants in the studied population. Vertical bar graphs depict the number of genes harboring deleterious rare variants within individual patients.

of L385P (Figure 3C). The RMSF mean and standard deviation of the WT and L385P were 0.14 ± 0.09 and 0.13 ± 0.08 nm, respectively (Figure 3D). However, the average RMSF was slightly decreased in L385P, and the overall change in the protein's flexibility was insignificant upon the mutation.

3.3.2. R_g and SASA Analyses. The R_g provides insight into the spatial arrangement of atoms around a central axis, with more compact structures exhibiting lower R_g values and less compact structures exhibiting higher R_g values. The R_g mean and standard deviation of the WT and L385P were 1.73 ± 0.01 and 1.77 ± 0.01 nm, respectively (Figure 4A,B). In these cases, the R_g value for L385P was higher, indicating that the protein has a more expanded or open conformation. R_g suggests that the L385P mutation caused an ESRRA protein to adopt a less compact structure than the WT, suggesting increased flexibility and less stability. The SASA computation involves assessing a protein's surface area that is accessible to the solvent and the buried surface area that is not accessible. We investigated the SASA of both the WT and L385P mutations of ESRRA to gain insights into their conformational dynamics throughout the simulation. The SASA mean and standard deviation of the WT and L385P were 109.02 ± 2.13 and 111.50 ± 2.52 nm^2, respectively (Figure 4C,D). L385P exhibited a slightly higher SASA value than the WT, confirming changes in the structure and increased solvent accessibility of the surface residues.

3.3.3. Intramolecular H-Bond and PCA Analyses. The stability of the protein's dynamic system was notably

influenced by hydrogen bonds (H-bonds), which play a significant role in determining the most stable conformation. The intramolecular H-bond mean and standard deviation of the WT and L385P were 150.68 ± 5.94 and 147.76 ± 6.62, respectively (Figure 5A,B). The WT has more intramolecular H-bonds than L385P, indicating that the WT is more stable or has a conformation different from that of the mutant. This conformation could potentially lead to differences in their function. PCA was employed to comprehensively understand dynamic properties based on MDS outcomes. The projection of the first two eigenvectors (PC1 vs PC2) for the WT and L385P forms was carried out to expose the movements of the atoms. The resulting projection in phase space revealed that L385P forms a different cluster than the WT, suggesting a change in the conformational space upon mutation. The plot in Figure 5C,D distinctly shows that L385P occupies a smaller space in phase space, whereas the WT occupies a larger space. This analysis sheds light on the dynamics and conformational changes in the proteins during the simulation.

3.3.4. R_g vs SASA Using a Kernel Density Estimation Joint Plot Analysis. Integrating analyses of both R_g and SASA offers a comprehensive view of a biomolecule's structural dynamics. The relationship between the R_g and SASA for L385P and WT ESRRA is graphically depicted by using a kernel density estimation joint plot (Figure 6). The most densely occupied conformations for the WT were identified at SASA 108 nm^2 and R_g 1.73 nm, whereas for the L385P mutant, they were

Table 1. Prediction Derived from PredictSNP Encompasses 53 Rare Variants across 23 Genes[a]

gene symbol	missense variant	existing variation	PredictSNP	MAPP	PhD-SNP	Poly-Phen-1	Poly-Phen-2	SIFT	SNAP	PAN-THER
ZNF717	V22A*	rs75138373	DEL	DEL	DEL	DEL	DEL	DEL	NEU	DEL
	V29M	rs73117241	DEL	NEU	DEL	DEL	DEL	DEL	NEU	DEL
	L39V*	rs879105149	DEL	DEL	DEL	DEL	DEL	DEL	NEU	DEL
	R350C	rs1962893	NEU	DEL	NEU	NEU	NEU	NEU	DEL	NEU
	L410V*	rs193921040	DEL	DEL	NEU	DEL	DEL	DEL	DEL	DEL
	T614I*	rs796390112	DEL	DEL	NEU	DEL	DEL	DEL	DEL	DEL
	T837K*	rs908906825	DEL	DEL	NEU	DEL	DEL	DEL	DEL	DEL
PSPH	R65H*	rs200442078	DEL	DEL	DEL	DEL	DEL	DEL	DEL	NEU
	L68P*	rs78067484	DEL	DEL	DEL	DEL	DEL	DEL	DEL	NEU
ESRRA	R376L	rs201971362	NEU	NEU	DEL	NEU	NEU	NEU	NEU	NEU
	L385P*	rs201072913	DEL	DEL	DEL	DEL	DEL	DEL	DEL	UNK
	L388F	rs79204587	DEL	DEL	DEL	DEL	DEL	DEL	NEU	UNK
	R389C	rs80310817	DEL	DEL	DEL	DEL	NEU	DEL	DEL	UNK
CEL	F80S	rs371426439	DEL	NEU	NEU	DEL	DEL	DEL	DEL	UNK
	I485T	rs77696629	NEU	NEU	NEU	NEU	NEU	DEL	NEU	UNK
IL16	E459K	rs200730945	DEL	NEU	DEL	DEL	DEL	NEU	UNK	-NA-
COL6A5	D442V	rs754422889	DEL	-NA-	DEL	DEL	DEL	DEL	DEL	UNK
	A581V	rs201277707	DEL	-NA-	DEL	NEU	NEU	DEL	DEL	UNK
	R1180C	rs543746446	DEL	-NA-	DEL	DEL	NEU	DEL	DEL	UNK
	T1884R	rs553758826	DEL	DEL	DEL	NEU	DEL	DEL	DEL	UNK
SEMA3G	R169W*	rs149481070	DEL	DEL	DEL	DEL	DEL	DEL	DEL	DEL
	P375S	rs766164256	DEL	NEU	DEL	DEL	DEL	DEL	DEL	UNK
	K385S	rs1271717265	NEU	NEU	NEU	NEU	NEU	DEL	NEU	UNK
ACAD10	P448H	rs370592166	DEL	NEU	DEL	DEL	DEL	DEL	NEU	UNK
ARRDC4	R391Q	rs575728067	NEU	NEU	DEL	NEU	DEL	NEU	NEU	UNK
COL15A1	K1001R	rs35544077	NEU	-NA-	NEU	DEL	DEL	NEU	NEU	UNK
	P1192S	rs142838918	NEU	NEU	NEU	NEU	NEU	NEU	NEU	UNK
SMPD4	F484V	rs148027738	DEL	DEL	DEL	NEU	DEL	DEL	DEL	NEU
CXCR1	F211L	rs142076386	NEU	DEL	NEU	NEU	DEL	DEL	NEU	UNK
PTPN7	G43D	rs751583629	DEL	NEU	DEL	DEL	DEL	DEL	DEL	UNK
	R76C*	rs372590798	DEL	NEU	DEL	DEL	DEL	DEL	DEL	DEL
	G160R	rs753470838	NEU	-NA-	NEU	NEU	DEL	DEL	NEU	UNK
DNMBP	A250T	rs368686292	DEL	DEL	DEL	NEU	DEL	NEU	DEL	NEU
	N1156S	rs147671401	NEU	NEU	NEU	NEU	DEL	DEL	DEL	NEU
	R1256C	rs1299200147	DEL	NEU	NEU	DEL	DEL	DEL	DEL	DEL
DRC7	T815M	rs115337501	DEL	NEU	NEU	DEL	DEL	DEL	NEU	UNK
CAVIN4	H79L*	rs775953195	DEL	DEL	DEL	DEL	DEL	DEL	DEL	DEL
	P228L	rs144516649	NEU	NEU	NEU	NEU	DEL	DEL	NEU	DEL
NLRX1	L679V	rs200846476	DEL	DEL	NEU	DEL	DEL	SIFT	NEU	UNK
MYBBP1A	R671W	rs139356015	DEL	NEU	NEU	DEL	DEL	DEL	NEU	DEL
	L883M*	rs375427927	DEL	DEL	NEU	DEL	DEL	DEL	DEL	DEL
CGN	R267C	rs575514202	NEU	NEU	NEU	NEU	NEU	DEL	NEU	UNK
	D959Y	rs769053104	DEL	DEL	NEU	DEL	DEL	DEL	NEU	UNK
	R1158C	rs200436542	DEL	DEL	NEU	DEL	DEL	DEL	DEL	UNK
ANKK1	T496N	rs201207634	NEU	NEU	NEU	NEU	NEU	DEL	DEL	NEU
SCART1	R184C	rs377020560	DEL	-NA-	DEL	DEL	DEL	DEL	DEL	UNK
	G627S	rs1057384230	DEL	DEL	DEL	NEU	DEL	DEL	DEL	UNK
	A835T	rs1051072711	NEU	DEL	DEL	NEU	NEU	NEU	NEU	UNK
ITLN2	A140V	rs137860437	DEL	DEL	NEU	DEL	DEL	DEL	NEU	UNK
	P289L	rs763748248	DEL	DEL	DEL	NEU	DEL	DEL	DEL	UNK
	D299Y	rs1475168897	DEL	DEL	DEL	DEL	NEU	DEL	DEL	UNK
IGSF10	Y2504C	rs561757162	DEL	DEL	DEL	DEL	DEL	DEL	UNK	-NA-
	D2614N	rs112889898	DEL	DEL	NEU	DEL	DEL	DEL	UNK	-NA-

[a]DEL—Deleterious; NEU—Neutral; UNK—Unknown; NA—not available; missense variant marked with an asterisk (*) indicate predictions of deleterious by seven or eight tools.

observed at SASA 111 nm^2 and R_g 1.77 nm. The results revealed that an elevation in the SASA, coupled with an augmentation in the R_g, signifies the potential unfolding or expansion of the molecule.

3.3.5. FEL and Secondary Structure Analyses. The Gibbs free energy changes were assessed by analyzing the FEL through two principal components for the WT and L385P to explore the conformational behavior. The deeper shades of

Table 2. Summarizes Prediction Scores from Align GVGD, ConSurf, and I-Mutant2.0 Analyses, Specifically for Variants Identified as Deleterious in PredictSNP[a]

gene symbol	missense variant	existing variation	align GVGD	ConSurf	I-Mutant2.0 DDG (kcal/mol)
ZNF717	V22A	rs75138373	class C55	8	−1.48
	L39V	rs879105149	class C25	9	−1.29
	L410V	rs193921040	class C25	7	−0.38
	T614I*	rs796390112	class C65	9	−1.07
	T837K	rs908906825	class C65	U	−1.54
PSPH	R65H	rs200442078	class C25	9	−1.52
	L68P	rs78067484	class C65	6	−0.27
ESRRA	L385P*	rs201072913	class C65	9	0.08
SEMA3G	R169W	rs149481070	class C65	6	−0.01
PTPN7	R76C	rs372590798	class C65	4	−0.98
CAVIN4	H79L*	rs775953195	class C65	9	−0.69
MYBBP1A	L883M	rs375427927	class C0	9	−0.51

[a]"U" signifies uncertainty; missense variant belonging to class C65 with a ConSurf score of 9 are marked with an asterisk (*).

Table 3. Displays the Prediction Scores from MutPred2 Analysis for Variants ZNF717 (T614I), ESRRA (L385P), and CAVIN4 (H79L)[a]

gene symbol	substitution	MutPred2 score	remarks
ZN717	T614I	0.044	NA
ESRRA	L385P	0.961	loss of helix (Pr = 0.29 | P = 0.01); altered stability (Pr = 0.13 | P = 0.03)
CAVIN4	H79L	0.707	altered disordered interface (Pr = 0.29 | P = 0.03); gain of helix (Pr = 0.27 | P = 0.05); altered DNA binding (Pr = 0.14 | P = 0.05)

[a]NA—not available.

blue in the plots indicate different conformational states with lower energy levels. The WT exhibits single basins (deeper shade of blue), indicating a more stable conformational state with lower energy, while L385P exhibits two distinct basins, indicating two different stable conformational states (Figure 7A,B). From this, we found that the WT has a dominant and highly stable conformation, unlike the more diverse and flexible conformational landscape with two stable states. The SSE analysis involves calculating and characterizing several structural elements based on the protein's three-dimensional structure, such as α-helices, β-sheets, turns, bends, and coil regions. We employed the DSSP classification method to determine the SSE for individual amino acids in both the L385P mutant and WT ESRRA proteins. The temporal evolution of these elements is illustrated in Figure 8. The segment of residues 392-402 (S1 region) in the WT exhibits a five-helix, undergoing a shift to turns and α-helix due to the L385P mutation. The segment of residues 282-298 (S2 region) in the WT exhibits β-sheets altered to bends and turns upon the L385P mutation. The segment of residues 242-256 (S3 region) in the WT exhibits α-helix, altered to bends and turns upon the L385P mutation. Overall, we found that the SSE values of L385P and WT ESRRA proteins exhibited negligible changes.

4. DISCUSSION

Mutations in the EGFR gene were recognized as oncogenic drivers in lung cancer, significantly altering the diagnosis and treatment. The EGFR mutations, found in about 10% of Caucasian and up to 50% of Asian NSCLC patients, exhibit a higher prevalence in women, nonsmokers, and those with adenocarcinoma.[36] In advanced NSCLC with EGFR mutations, TKIs like erlotinib, gefitinib, and afatinib initially offer prolonged PFS compared to chemotherapy but resistance often emerges within a year. Osimertinib is a second-line option for T790M-positive cases, and intercalated regimens show extended survival. Accurately assessing a patient's likelihood of progressing on TKI therapy remains challenging but is vital for personalized treatment and optional outcomes.[37] An earlier study found that EGFR mutant lung SCC patients responded less favorably to EGFR-TKI treatment than to EGFR mutant lung adenocarcinoma.

In lung SCC, the concurrent presence of EGFR mutations along with mutations in the CREBBP, ZNF217, or Wnt pathway was found to correlate negatively with PFS.[38] An earlier study suggests that the occurrence of the EGFR T790M mutation in patients experiencing disease progression on TKIs is probably higher than previously thought. This disease progression is especially evident when considering mutations happening at the AF. They found that Osimertinib treatment proved advantageous for patients experiencing disease progression with the low frequency of EGFR T790M mutation.[19] WES is a powerful tool that can be used to identify deleterious rare missense variants. It involves the targeted sequencing of exonic regions of all known protein-coding genes.[39] Our study is the first to examine the profiles of deleterious rare variants in NSCLC patients with EGFR mutations who have become resistant to erlotinib-TKI. We initially acquired WES data of 16 NSCLC-TKI-resistant samples from the SRA database. Subsequent analysis led to the identification of 41.8% of novel variants and 68.6% of existing variants in samples. Following this, we refined our analysis by filtering for missense variants with an MAF of less than 1%. We specifically focused on variants deemed deleterious by the SIFT and PolyPhen-2 tools. This process led to the identification of 53 rare deleterious variants across 23 genes. Subsequently, we subjected the 53 rare variants to Insilco pathogenicity assessment tools including PredictSNP, Align GVGD, Consurf, I-Mutant2.0, and Mutpred2. PredictSNP analysis resulted in 12 deleterious rare variants in 7 genes, namely ZNF717, PSPH, ESRRA, SEMA3G, PTPN7, CAVIN4, and MYBBP1A. Interestingly, the Insilco pathogenicity analysis of all the tools highlighted ESRRA L385P as notably more deleterious than the other variants.

ZNF717, also known as OBI1, is a possible E3 ubiquitin ligase that facilitates the initiation of replication by encouraging
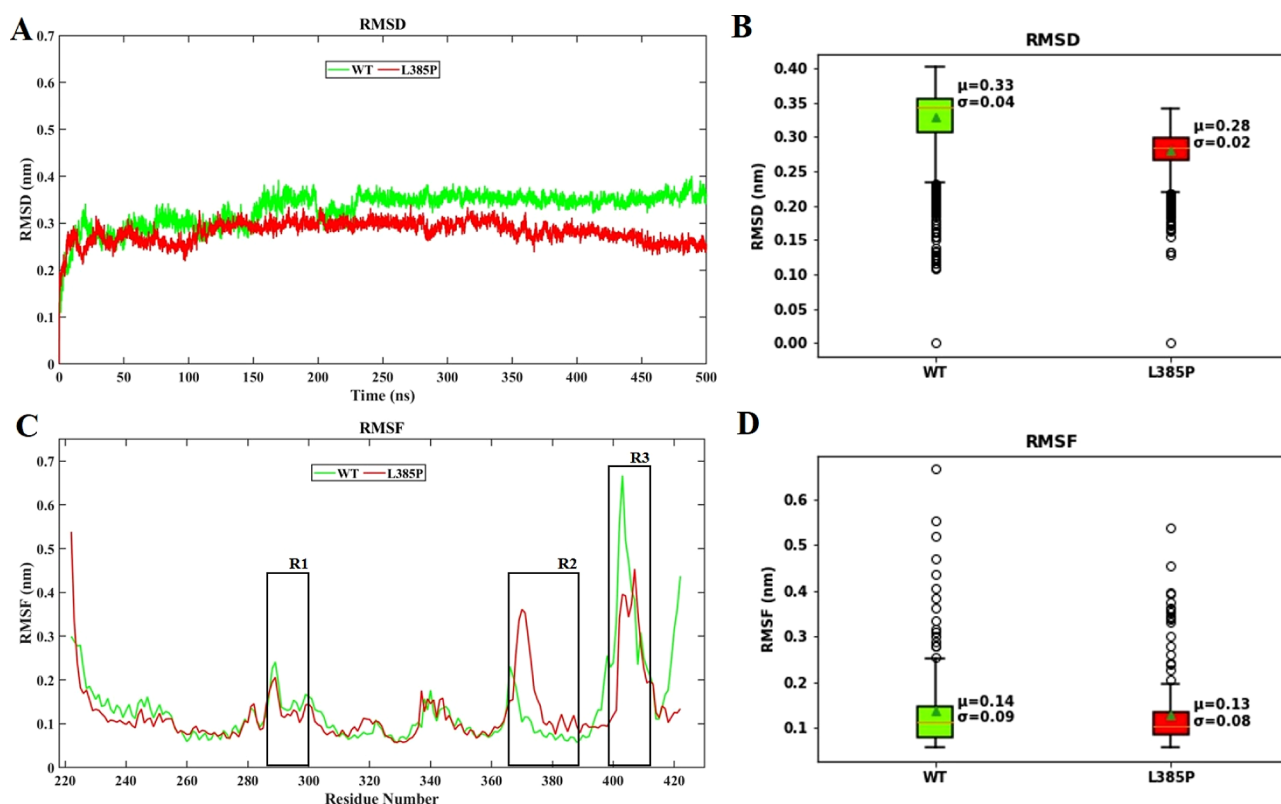
Figure 3. (A) rmsd analysis compares the ESRRA WT and L385P mutation over 500 ns MDS, with time (ns) on the X-axis and rmsd values (nm) on the Y-axis. (B) Box plot illustrates the rmsd analysis of the ESRRA WT and the L385P mutation, presenting key statistical measures such as the mean, median, and standard deviation. (c) RMSF analysis of the ESRRA WT and L385P mutation throughout 500 ns MDS, with residue number on the X-axis and RMSF values (nm) on the Y-axis. (D) Box plot illustrates the RMSF analysis of the ESRRA WT and the L385P mutation.

the firing of replication origins. Silencing *OBI1* leads to impaired replication origin firing.[40] Moreover, *ZNF717* plays a crucial role in the osteogenic differentiation of mesenchymal stem cells.[41] Recently, variations in *ZNF717* have been identified in individuals diagnosed with lung adenocarcinoma, familial lung cancer, colorectal cancer, and gastric cancer.[42−45]

Phosphoserine phosphatase (PSPH) plays a pivotal role as a key enzyme in the L-serine synthesis pathway. PSPH significantly influences patient prognosis and controls lung cancer, cell invasion, and colony farming. Based on recent data, using PSPH's particular interaction and dephosphorylation activity could result in novel therapeutic strategies for lung cancer treatment.[46]

Semaphorin 3G (*SEMA3G*) is a type of class 3 secreted semaphorins primarily found in endothelial cells.[47] *SEMA3G* has reportedly been implicated in inhibiting migration and invasion in the case of glioma.[48] According to research, *SEMA3G* exhibits expression in organs rich in the vasculature, with the lung demonstrating the highest expression levels, followed by the kidney, heart, and ovary.[49] *SEMA3G* is present in the migratory path of GnRH neurons and is expressed during the development of the pituitary gland. A mutated form of *SEMA3G* modifies its binding properties to PlexinA receptors both in computational and laboratory experiments, reducing its impacts on the migration of immortalized GnRH neurons.[50]

Protein tyrosine phosphate nonreceptor type 7 (*PTPN7*) plays an anti-inflammatory role by inhibiting Erk1/2 and p38, thereby reducing the pro-inflammatory cytokine tumor necrosis factor production.[51] Research indicates that *PTPN7*

expression correlates with immune features in breast cancer, suggesting its potential as a predictive biomarker for immunotherapy across various cancer types.[52]

The *cavin4a* gene produces cavin-4, also called the muscle-restricted coiled-coil (MURC) protein, a member of the cavin family. Cavin proteins play a crucial role in the structure and function of caveolae, cellular membrane invaginations.[53] *CAVIN4* expression rises during muscle regeneration in response to injury by activating the extracellular signal-regulated kinase pathway.[54] Cavin-4 is found in limited amounts in various cell types, including embryonic fibroblasts, where it can interact with Cavin-2. Cavin-1 and Cavin-2 play dual roles in cancer, acting as tumor suppressors or promoters depending on the cancer type and stage.[55]

MYB binding protein 1A (MYBBP1A or p10) is a corepressor for various transcription factors engaged in numerous physiological processes, thereby functioning as a tumor suppressor with significant relevance to various aspects of cell physiology crucial for tumorigenesis. *MYBBP1A* undergoes frequent loss of heterozygosity (50%−80%) in diverse cancers like breast, ovarian, astrocytomas, bladder, medulloblastomas, osteosarcomas, leukemias, and lung cancers. This loss indicates the likely presence of tumor suppressor genes in this region.[56] A prior study found that 24-mCAF suppresses the activity of AKT and aurora B kinases, and Ser/ Thr kinases crucial for *MYBBP1A* regulation and significant targets in NSCLC.[57]

ESRRA or ERRα, a nuclear receptor, is highly expressed in metabolically active organs and functions as a transcription factor. Its elevated expression is observed in various
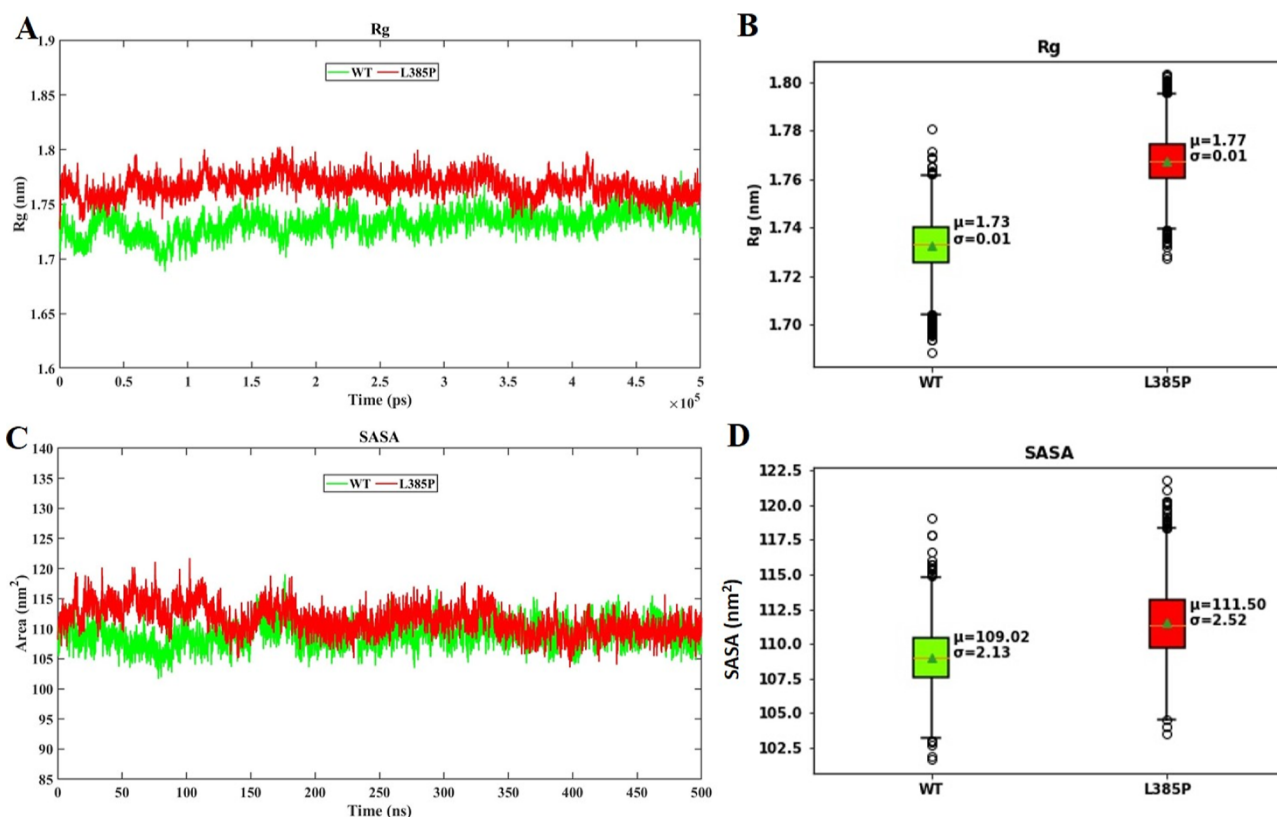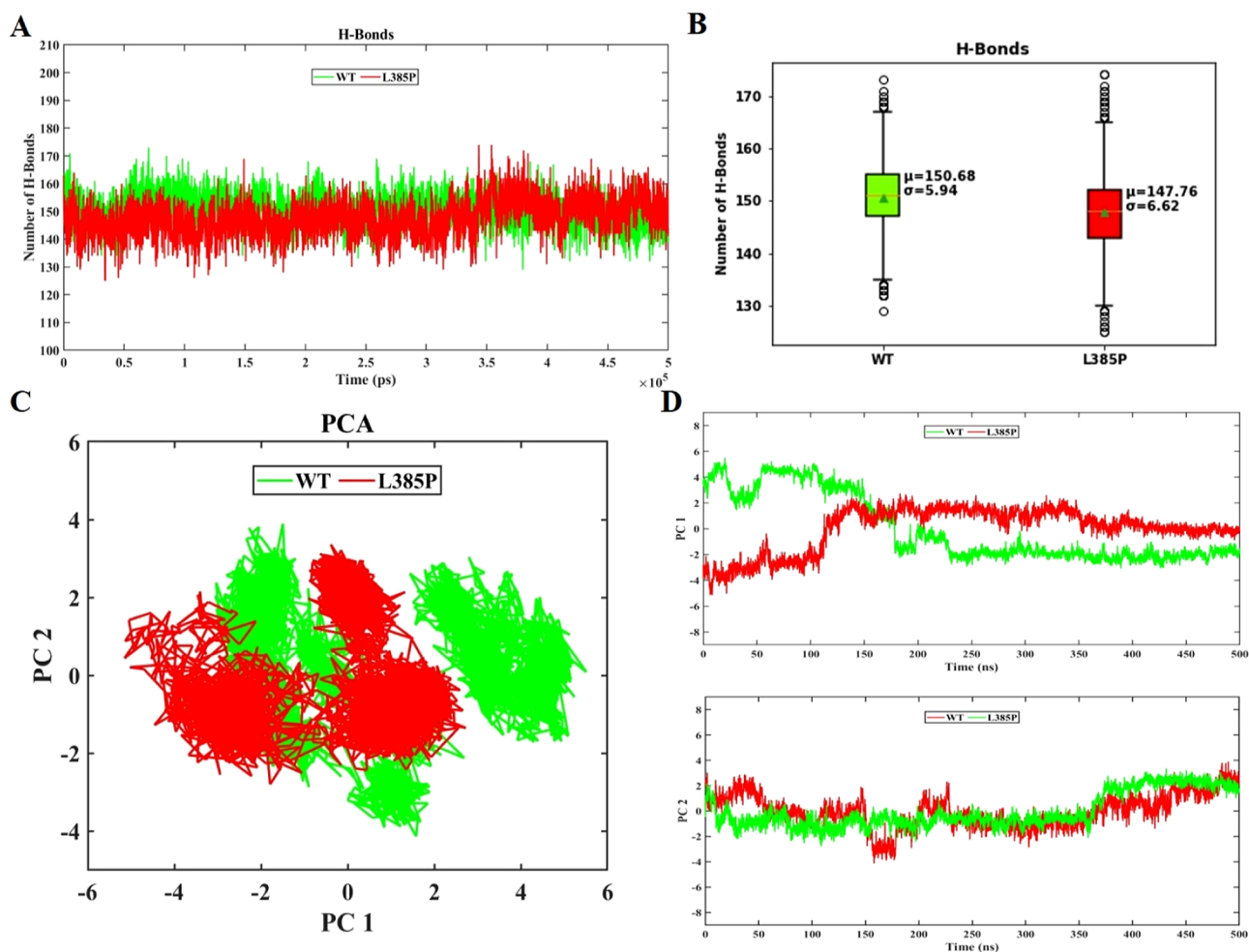
Figure 4. (A) The R_g analysis compares the ESRRA WT and L385P mutation over 500 ns MDS, with time (ps) on the X-axis and R_g values (nm) on the Y-axis. (B) Box plot illustrates the R_g analysis of the ESRRA WT and the L385P mutation, presenting key statistical measures such as the mean, median, and standard deviation. (c) SASA analysis of the ESRRA WT and L385P mutation throughout 500 ns MDS, with time (ns) on the X-axis and area values (nm^2) on the Y-axis. (D) Box plot illustrates the SASA analysis of the ESRRA WT and the L385P mutation.

malignancies. Silencing *ESRRA* hindered gastric cancer cell proliferation, viability, migration, and invasion. *ESRRA* down-regulation induced G2M cycle arrest by suppressing the *CDC25C/CDK1/CyclinB1* pathway through the regulation of *DSN1* expression, as demonstrated by flow cytometry and western blot analysis.[58] ERRα is typically expressed in the lungs and activates nuclear genes without ligand binding, like estrogen. In NSCLC, ERRα expression is notably elevated compared with healthy individuals. The well-established fact is that ERα and ERβ share 93% identity in the DNA-binding domain and 60% identity in the ligand-binding domain. ERRα enhances proliferation, invasion, and migration by suppressing tumor suppressor proteins p53 and pRB. Additionally, ERRα expedites the G2-M transition in cell division. The activation of NSCLC metastasis is primarily facilitated by epithelial-to-mesenchymal transition (EMT) and slug induction, a transcription factor associated with EMT.[59] In our investigation, we discovered the *ESRRA* L385P mutation in 7 out of 16 (43.75%) EGFR-TKI-resistant cases of NSCLC, predominantly in males and individuals who harbored *EGFR* exon 19 deletions and showed resistance to TKI therapy.

MDS serves as a valuable tool in uncovering the structural impact of mutations on proteins, elucidating alterations in stability and interactions with other molecules. This capability provides crucial insights into the molecular mechanisms associated with genetic disease, offering a foundation for the potential development of targeted therapeutic strategies.[60] In our investigation, we explored the changes in the L385P ESRRA variant using MDS, closely examining its deviation from the WT configuration under physiological conditions.

The simulations, extended over 500 ns, comprehensively evaluated structural aspects including stability, flexibility, hydrogen bonding, and solvation dynamics of the mutated protein.

rmsd quantifies the average structural difference between a set of backbone atoms, commonly applied to assess molecular conformation changes from a reference structure. We obtained a lower rmsd in L385P than in the WT, which indicates that mutation reduced the ESRRA flexibility. RMSF measures protein flexibility by calculating the average deviation of each atom in a protein structure from its mean position throughout an MDS. Following rmsd, L385P showed a slightly reduced average RMSF compared to that of the WT, reflecting that this mutation decreases the flexibility of the ESRRA protein. R_g is a crucial parameter for characterizing the equilibrium conformation of a system, particularly in protein analysis. It reflects the protein structure level of compression and indicates whether the polypeptide chain is folded or unfolded.[61] In our observations, we noted that an elevation in the R_g value in L385P indicates a decrease in protein compactness, which is frequently linked to unfolding or conformational changes in the protein structure. SASA measures an amino acid's interaction with its environment, including the solvent and the protein environment.[62] Variations in the SASA plot, whether an increase or decrease, indicate alterations in the involved amino acid residues, consequently influencing the tertiary structure of a protein. The increased SASA value for L385P indicates that the protein has been exposed to more solvents, suggesting that the protein is more relaxed and flexible. H-bonds are important in stabilizing proteins.[63] L385P

Figure 5. (A) Intramolecular H-bond analysis compares the ESRRA WT and L385P mutation over 500 ns MDS, with time (ps) on the *X*-axis and the number of H-bonds on the *Y*-axis. (B) Box plot illustrates the intramolecular H-bond analysis of the ESRRA WT and the L385P mutation, presenting key statistical measures such as mean, median, and standard deviation. (C) PCA analysis of the ESRRA WT and L385P mutation, with the PC1 (nm) on the *X*-axis and PC2 (nm) on the *Y*-axis. (D) PC1 and PC2 analyses of the ESRRA WT and L385P mutation throughout 500 ns of MDS.
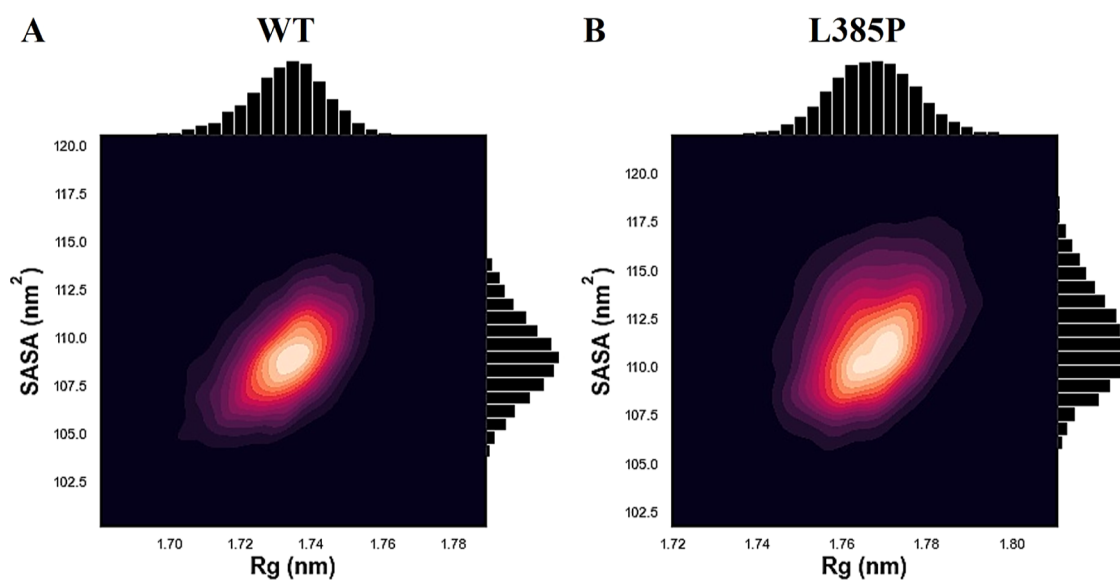
Figure 6. Kernel density estimation joint plot was utilized to depict the conformational dynamics of the (A) ESRRA WT and (B) L385P mutation throughout the MDS. The *X*-axis represents R_g (nm), while the *Y*-axis illustrates the SASA (nm^2).
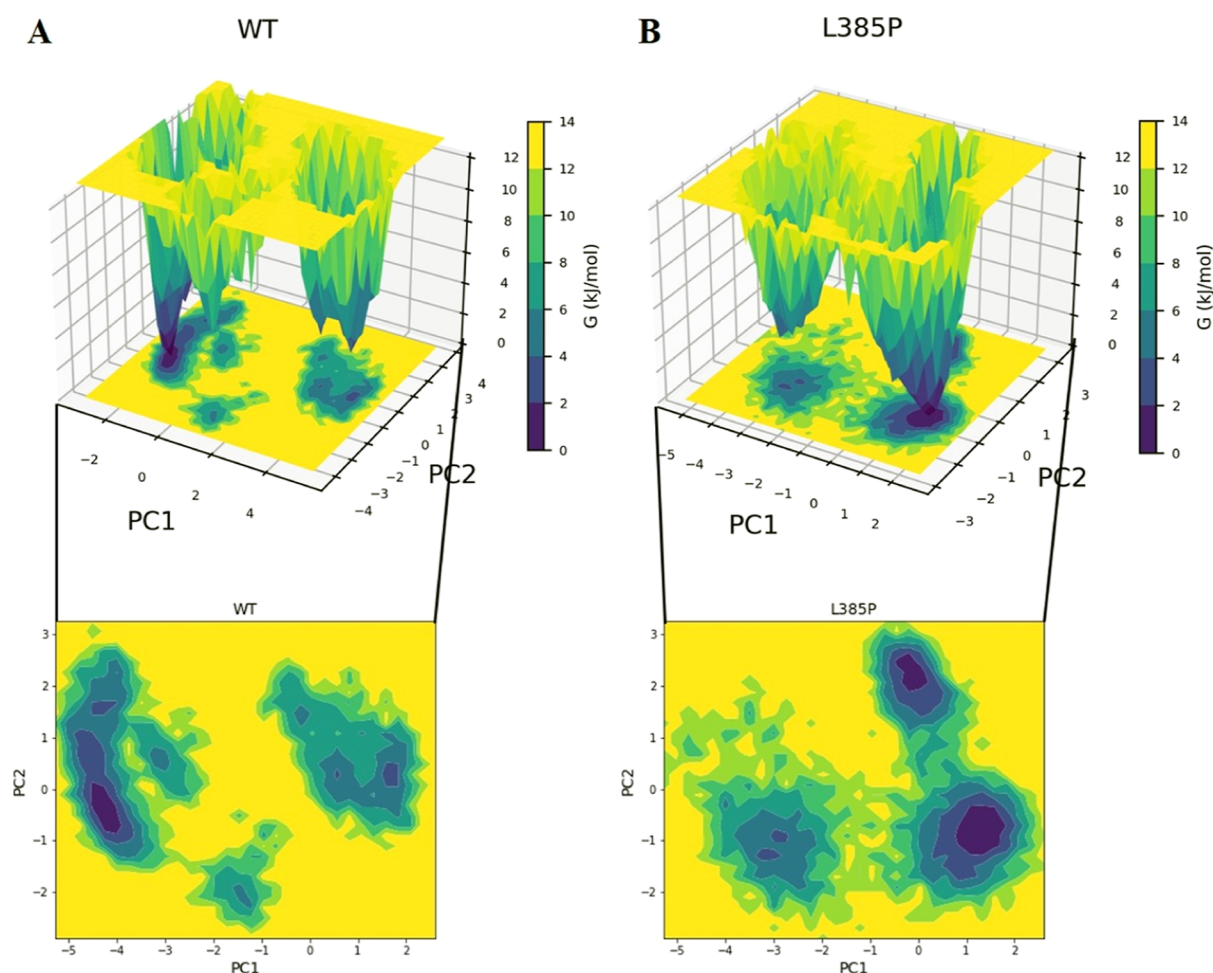
Figure 7. Two-dimensional contour and the three-dimensional FEL for the (A) ESRRA WT and (B) L385P mutation based on 500 ns of MDS. Color gradients graphically show the free energy values expressed in kJ/mol.

has lower average H-bonds than the WT, suggesting that this mutation may result in a less compact and more flexible ESRRA protein structure. PCA analysis of the protein conformational space reveals that L385P occupies a smaller phase space, implying a more limited conformational range.[64] The WT explores a broader phase space, indicating a more flexible conformational space. This analysis suggests that the L385P mutation imposes restrictions or constraints on the protein's conformational dynamics compared to the WT. The comparison between R_g and SASA results reveals that the L385P mutation leads to an increase in SASA and a rise in R_g, indicating potential unfolding or expansion of the molecule (Figure 6). FEL analysis discovered that the WT exhibits a predominant and exceptionally stable conformation, while the L385P variant demonstrates a more varied and flexible landscape with two stable states (Figure 7). The FEL analysis revealed discrete energy minima corresponding to different conformational states.[65] The mutant L385P showed different energy minima than the WT. According to the SSE analysis,[66] L385P and the WT showed insignificant changes in the overall structure despite minor deviations in the elements.

According to a recent study, resistant NSCLC cells treated with gefitinib or Osimertinib had synchronized EGFR reactivation and ERRα re-expression. Extended exposure to TKIs (gefitinib or Osimertinib) induces the accumulation of cholesterol in lipid fats in NSCLC. This accumulation creates a platform for the interaction of EGFR and Src, leading to the reactivation of the *EGFR/Src/Erk* signaling pathway. The level of nuclear translocation of *SP1* increases, enhancing *ERRα* transcription by binding to its promoter. The re-expression of ERRα protein contributes to a context of reactive oxygen species detoxification, facilitating cell survival despite continuous exposure to EGFR-TKIs. Their findings also showed that pharmacologically reducing cholesterol levels and inhibiting ERRα restored resistant cells and in vivo xenograft tumors.[67] The identified L385P mutation in our investigation resides within the ligand-binding domain of ERRα.[68] This mutation can potentially cause structural unfolding in ERRα proteins, which may influence their function and lead to changes in ERRα expression. Overall, our MDS study showed that ESRRA L385P increased the flexibility or unfolding of the protein structure compared with the WT. This in-depth MDS analysis provides valuable insights into the potential functional consequences of the L385P mutation in ESRRA, elucidating its structural behavior and assisting in interpreting its biological significance. The study's limitations include the absence of normal tissue samples for comparison, making it difficult to determine if identified missense variants are tumor-specific. Moreover, the exclusive focus on erlotinib-resistant cases may lack clinical relevance, considering osimertinib as the standard for EGFR mutant NSCLC. Caution is recommended in interpreting the results, highlighting the necessity for future

WT
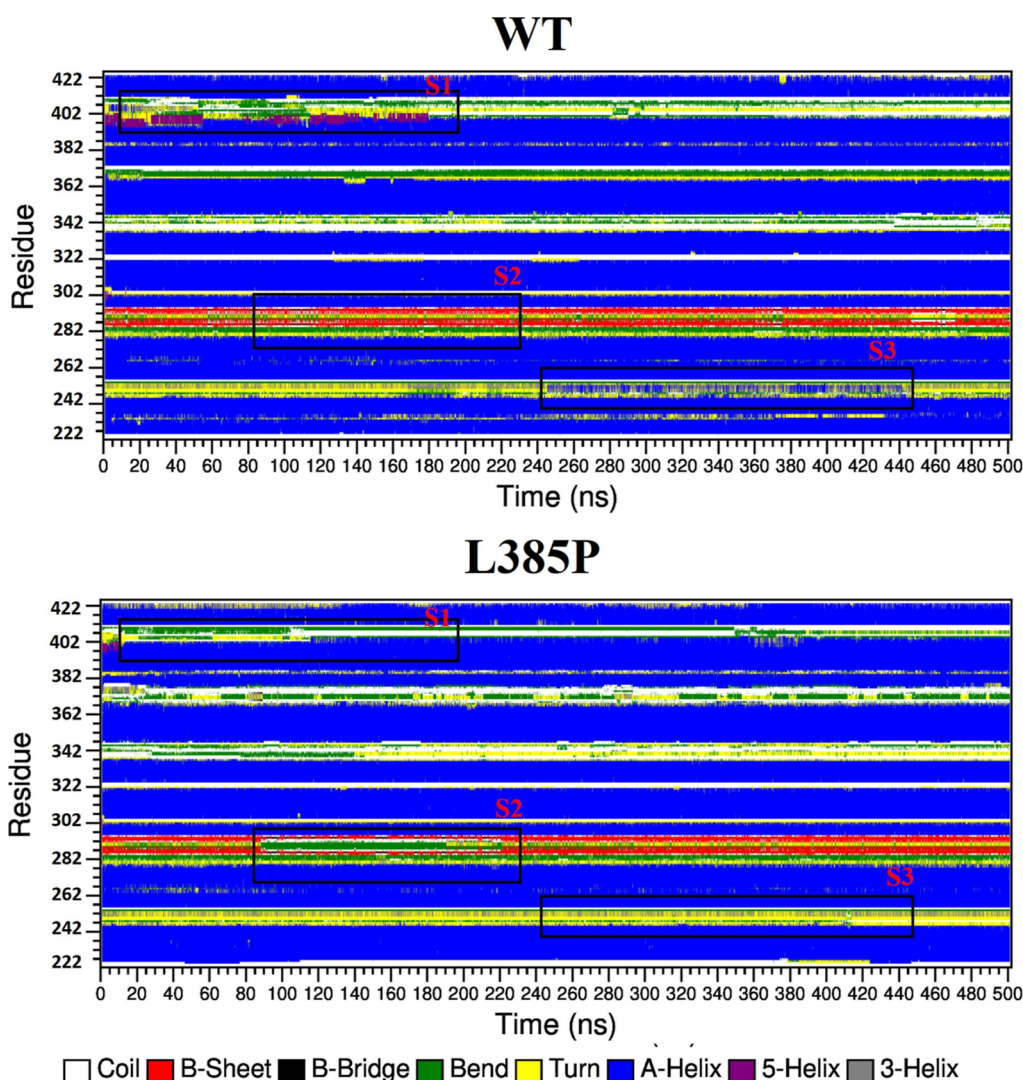
L385P

Figure 8. DSSP secondary structure analysis was conducted, tracking the evolution over simulation time for the trajectories of both the ESRRA (A) WT and (B) L385P mutation. The color-coded representation assigns a specific secondary structure type to each residue as the simulation progresses: White for the random coil, red for the β-sheet, black for the β-bridge, green for the bend, yellow for the turn, blue for the α-helix, purple for the 5-helix, and gray for the 3-helix.

studies with experimental validation, diverse samples, and alignment with current clinical practices to enhance our understanding of these identified missense variants.

5. CONCLUSIONS

Our study uncovered 53 deleterious rare missense variants across 23 genes by employing WES in EGFR-TKI erlotinib-resistant patients. These variants exhibited an MAF of less than 1% and were predicted to be deleterious according to the SIFT and PolyPhen-2 tools. Further scrutinizing with pathogenic tools such as PredictSNP, 12 rare missense variants were found in 7 genes. Our data suggest that rare missense variants in the *ZNF717, PSPH, ESRRA, SEMA3G, PTPN7, CAVIN4*, and *MYBBP1A* genes have a substantial role in NSCLC, urging for additional research even though the tumorigenic mechanism of these variants is yet unknown. The target genes identified in our study can serve as valuable biomarkers for detecting and diagnosing NSCLC. Furthermore, Align GVGD, ConSurf, and I-Mutant2.0 assessments indicated that the ESRRA L385P mutation exhibited a higher degree of deleterious impact than other mutations. The comprehensive insights gained from

MDS enhance our understanding of the structural and dynamic consequences of the L385P ESRRA mutation, providing valuable implications for subsequent therapeutic considerations and targeted interventions.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.3c10229.

> Overview of the study workflow, details of NSCLC with EGFR-TKI-resistant samples obtained from the SRA-NCBI database, total SNVs and variant distribution in each sample, and heatmap illustrating 53 deleterious rare missense variants across 23 genes in the WES of 16 NSCLC-resistant EGFR-TKI patients (PDF)

■ AUTHOR INFORMATION

Corresponding Author

George Priya C Doss − *Laboratory of Integrative Genomics, Department of Integrative Biology, School of BioSciences and*

Technology, Vellore Institute of Technology, Vellore, Tamil Nadu 632014, India; ● orcid.org/0000-0002-5971-8290; Email: georgepriyadoss@vit.ac.in

Author

Ambritha Balasundaram − *Laboratory of Integrative Genomics, Department of Integrative Biology, School of BioSciences and Technology, Vellore Institute of Technology, Vellore, Tamil Nadu 632014, India*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acsomega.3c10229

Author Contributions

A.B. and C.G.P.D. were involved in the study's design. A.B. was involved in the data collection and experimentation. A.B. was involved in acquiring, analyzing, and interpreting the results. A.B. drafted the manuscript. C.G.P.D. supervised the entire study and was involved in study design, acquiring, analyzing, and understanding the data, and critically reviewing the manuscript. All authors edited and approved the submitted version of the article.

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Sung, H.; Ferlay, J.; Siegel, R. L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *Ca-Cancer J. Clin.* **2021**, *71* (3), 209−249.

(2) Inamura, K. Lung Cancer: Understanding Its Molecular Pathology and the 2015 WHO Classification. *Front. Oncol.* **2017**, *7*, 193.

(3) Kris, M. G.; Johnson, B. E.; Berry, L. D.; Kwiatkowski, D. J.; Iafrate, A. J.; Wistuba, I. I.; Varella-Garcia, M.; Franklin, W. A.; Aronson, S. L.; Su, P.-F.; Shyr, Y.; Camidge, D. R.; Sequist, L. V.; Glisson, B. S.; Khuri, F. R.; Garon, E. B.; Pao, W.; Rudin, C.; Schiller, J.; Haura, E. B.; Socinski, M.; Shirai, K.; Chen, H.; Giaccone, G.; Ladanyi, M.; Kugler, K.; Minna, J. D.; Bunn, P. A. Using Multiplexed Assays of Oncogenic Drivers in Lung Cancers to Select Targeted Drugs. *JAMA* **2014**, *311* (19), 1998−2006.

(4) Cagle, P. T.; Raparia, K.; Portier, B. P. Emerging Biomarkers in Personalized Therapy of Lung Cancer. *Adv. Exp. Med. Biol.* **2016**, *890*, 25−36.

(5) Ellis, L. M.; Hicklin, D. J. Resistance to Targeted Therapies: Refining Anticancer Therapy in the Era of Molecular Oncology. *Clin. Cancer Res.* **2009**, *15* (24), 7471−7478.

(6) Rodak, O.; Peris-Díaz, M. D.; Olbromski, M.; Podhorska-Okołów, M.; Dzięgiel, P. Current Landscape of Non-Small Cell Lung Cancer: Epidemiology, Histological Classification, Targeted Therapies, and Immunotherapy. *Cancers* **2021**, *13* (18), 4705.

(7) Giuliani, J.; Martelli, S.; Remo, A.; Bonetti, A. Primary TKI Resistance in Advanced Non-Small Cell Lung Cancer with EGFR Mutation: An Open Question. *Tumori* **2015**, *101* (4), e115−e117.

(8) Riely, G. J.; Pao, W.; Pham, D.; Li, A. R.; Rizvi, N.; Venkatraman, E. S.; Zakowski, M. F.; Kris, M. G.; Ladanyi, M.; Miller, V. A. Clinical Course of Patients with Non-Small Cell Lung Cancer and Epidermal Growth Factor Receptor Exon 19 and Exon 21 Mutations Treated with Gefitinib or Erlotinib. *Clin. Cancer Res.* **2006**, *12* (3), 839−844.

(9) Guardiola, S.; Varese, M.; Sánchez-Navarro, M.; Giralt, E. A Third Shot at EGFR: New Opportunities in Cancer Therapy. *Trends Pharmacol. Sci.* **2019**, *40* (12), 941−955.

(10) Chmielecki, J.; Gray, J. E.; Cheng, Y.; Ohe, Y.; Imamura, F.; Cho, B. C.; Lin, M.-C.; Majem, M.; Shah, R.; Rukazenkov, Y.; Todd, A.; Markovets, A.; Barrett, J. C.; Hartmaier, R. J.; Ramalingam, S. S. Candidate Mechanisms of Acquired Resistance to First-Line Osimertinib in EGFR-Mutated Advanced Non-Small Cell Lung Cancer. *Nat. Commun.* **2023**, *14* (1), 1070.

(11) Lin, C.-C.; Shih, J.-Y.; Yu, C.-J.; Ho, C.-C.; Liao, W.-Y.; Lee, J.-H.; Tsai, T.-H.; Su, K.-Y.; Hsieh, M.-S.; Chang, Y.-L.; Bai, Y.-Y.; Huang, D. D.-R.; Thress, K. S.; Yang, J. C.-H. Outcomes in Patients with Non-Small-Cell Lung Cancer and Acquired Thr790Met Mutation Treated with Osimertinib: A Genomic Study. *Lancet Respir. Med.* **2018**, *6* (2), 107−116.

(12) Oxnard, G. R.; Hu, Y.; Mileham, K. F.; Husain, H.; Costa, D. B.; Tracy, P.; Feeney, N.; Sholl, L. M.; Dahlberg, S. E.; Redig, A. J.; Kwiatkowski, D. J.; Rabin, M. S.; Paweletz, C. P.; Thress, K. S.; Jänne, P. A. Assessment of Resistance Mechanisms and Clinical Implications in Patients With EGFR T790M-Positive Lung Cancer and Acquired Resistance to Osimertinib. *JAMA Oncol.* **2018**, *4* (11), 1527−1534.

(13) Oxnard, G.; Hu, Y.; Mileham, K.; Tracy, P.; Feeney, N.; Sholl, L.; Paweletz, C.; Thress, K.; Jänne, P. OA 09.02 Osimertinib Resistance Mediated by Loss of EGFR T790M Is Associated with Early Resistance and Competing Resistance Mechanisms. *J. Thorac. Oncol.* **2017**, *12* (11), S1767−S1768.

(14) van der−>Wekken, A.; Saber, A.; Hiltermann, T. J. N.; Kok, K.; van den−>Berg, A.; Groen, H. J. M. Resistance Mechanisms after Tyrosine Kinase Inhibitors Afatinib and Crizotinib in Non-Small Cell Lung Cancer, a Review of the Literature. *Crit. Rev. Oncol. Hematol.* **2016**, *100*, 107−116.

(15) Passaro, A.; Jänne, P. A.; Mok, T.; Peters, S. Overcoming Therapy Resistance in EGFR-Mutant Lung Cancer. *Nat. Cancer* **2021**, *2* (4), 377−391.

(16) Nair, S. V.; Madhulaxmi; Thomas, G.; Ankathil, R. Next-Generation Sequencing in Cancer. *J. Maxillofac. Oral Surg.* **2021**, *20* (3), 340−344.

(17) Bartha, A. ́.; Győrffy, B. Comprehensive Outline of Whole Exome Sequencing Data Analysis Tools Available in Clinical Oncology. *Cancers* **2019**, *11* (11), 1725.

(18) GB Editorial Team. Closure of the NCBI SRA and Implications for the Long-Term Future of Genomics Data Storage. *Genome Biol.* **2011**, *12* (3), 402.

(19) Joshi, A.; Butle, A.; Hait, S.; Mishra, R.; Trivedi, V.; Thorat, R.; Choughule, A.; Noronha, V.; Prabhash, K.; Dutt, A. Osimertinib for Lung Cancer Cells Harboring Low-Frequency EGFR T790M Mutation. *Transl. Oncol.* **2022**, *22*, 101461.

(20) Leinonen, R.; Sugawara, H.; Shumway, M. The Sequence Read Archive. *Nucleic Acids Res.* **2011**, *39* (Database), D19−D21.

(21) Bolger, A. M.; Lohse, M.; Usadel, B. Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* **2014**, *30* (15), 2114−2120.

(22) Li, H.; Durbin, R. Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform. *Bioinformatics* **2009**, *25* (14), 1754−1760.

(23) Danecek, P.; Bonfield, J. K.; Liddle, J.; Marshall, J.; Ohan, V.; Pollard, M. O.; Whitwham, A.; Keane, T.; McCarthy, S. A.; Davies, R. M.; Li, H. Twelve Years of SAMtools and BCFtools. *GigaScience* **2021**, *10* (2), giab008.

(24) Hunt, S. E.; Moore, B.; Amode, R. M.; Armean, I. M.; Lemos, D.; Mushtaq, A.; Parton, A.; Schuilenburg, H.; Szpak, M.; Thormann, A.; Perry, E.; Trevanion, S. J.; Flicek, P.; Yates, A. D.; Cunningham, F. Annotating and Prioritizing Genomic Variants Using the Ensembl Variant Effect Predictor—A Tutorial. *Hum. Mutat.* **2022**, *43* (8), 986−997.

(25) Bendl, J.; Stourac, J.; Salanda, O.; Pavelka, A.; Wieben, E. D.; Zendulka, J.; Brezovsky, J.; Damborsky, J. PredictSNP: Robust and Accurate Consensus Classifier for Prediction of Disease-Related Mutations. *PLoS Comput. Biol.* **2014**, *10* (1), No. e1003440.

(26) Tavtigian, S. V.; Deffenbaugh, A. M.; Yin, L.; Judkins, T.; Scholl, T.; Samollow, P. B.; Silva, D. de; Zharkikh, A.; Thomas, A. Comprehensive Statistical Study of 452 BRCA1Missense Substitutions with Classification of Eight Recurrent Substitutions as Neutral. *J. Med. Genet.* **2006**, *43* (4), 295−305.

(27) Mathe, E.; Olivier, M.; Kato, S.; Ishioka, C.; Hainaut, P.; Tavtigian, S. V. Computational Approaches for Predicting the Biological Effect of P53 Missense Mutations: A Comparison of Three Sequence Analysis Based Methods. *Nucleic Acids Res.* **2006**, *34* (5), 1317−1325.

(28) Ashkenazy, H.; Abadi, S.; Martz, E.; Chay, O.; Mayrose, I.; Pupko, T.; Ben-Tal, N. ConSurf 2016: An Improved Methodology to Estimate and Visualize Evolutionary Conservation in Macromolecules. *Nucleic Acids Res.* **2016**, *44* (W1), W344−W350.

(29) Capriotti, E.; Fariselli, P.; Casadio, R. I-Mutant2.0: Predicting Stability Changes upon Mutation from the Protein Sequence or Structure. *Nucleic Acids Res.* **2005**, *33* (Web Server), W306−W310.

(30) Li, B.; Krishnan, V. G.; Mort, M. E.; Xin, F.; Kamati, K. K.; Cooper, D. N.; Mooney, S. D.; Radivojac, P. Automated Inference of Molecular Mechanisms of Disease from Amino Acid Substitutions. *Bioinformatics* **2009**, *25* (21), 2744−2750.

(31) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1−2*, 19−25.

(32) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. CHARMM General Force Field: A Force Field for Drug-like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.* **2010**, *31* (4), 671−690.

(33) Patch, R. J.; Searle, L. L.; Kim, A. J.; De, D.; Zhu, X.; Askari, H. B.; O'Neill, J. C.; Abad, M. C.; Rentzeperis, D.; Liu, J.; Kemmerer, M.; Lin, L.; Kasturi, J.; Geisler, J. G.; Lenhard, J. M.; Player, M. R.; Gaul, M. D. Identification of Diaryl Ether-Based Ligands for Estrogen-Related Receptor *α* as Potential Antidiabetic Agents. *J. Med. Chem.* **2011**, *54* (3), 788−808.

(34) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18* (12), 1463−1472.

(35) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103* (19), 8577−8593.

(36) Kuiper, J. L.; Hashemi, S. M. S.; Thunnissen, E.; Snijders, P. J. F.; Grünberg, K.; Bloemena, E.; Sie, D.; Postmus, P. E.; Heideman, D. a. M.; Smit, E. F. Non-Classic EGFR Mutations in a Cohort of Dutch EGFR-Mutated NSCLC Patients and Outcomes Following EGFR-TKI Treatment. *Br. J. Cancer* **2016**, *115* (12), 1504−1512.

(37) Song, J.; Shi, J.; Dong, D.; Fang, M.; Zhong, W.; Wang, K.; Wu, N.; Huang, Y.; Liu, Z.; Cheng, Y.; Gan, Y.; Zhou, Y.; Zhou, P.; Chen, B.; Liang, C.; Liu, Z.; Li, W.; Tian, J. A New Approach to Predict Progression-Free Survival in Stage IV EGFR-Mutant NSCLC Patients with EGFR-TKI Therapy. *Clin. Cancer Res.* **2018**, *24* (15), 3583−3592.

(38) Jin, R.; Peng, L.; Shou, J.; Wang, J.; Jin, Y.; Liang, F.; Zhao, J.; Wu, M.; Li, Q.; Zhang, B.; Wu, X.; Lan, F.; Xia, L.; Yan, J.; Shao, Y.; Stebbing, J.; Shen, H.; Li, W.; Xia, Y. EGFR-Mutated Squamous Cell Lung Cancer and Its Association With Outcomes. *Front. Oncol.* **2021**, *11*, 680804.

(39) Tetreault, M.; Bareke, E.; Nadaf, J.; Alirezaie, N.; Majewski, J. Whole-Exome Sequencing as a Diagnostic Tool: Current Challenges and Future Opportunities. *Expert Rev. Mol. Diagn.* **2015**, *15* (6), 749−760.

(40) Coulombe, P.; Nassar, J.; Peiffer, I.; Stanojcic, S.; Sterkers, Y.; Delamarre, A.; Bocquet, S.; Méchali, M. The ORC Ubiquitin Ligase OBI1 Promotes DNA Replication Origin Firing. *Nat. Commun.* **2019**, *10* (1), 2426.

(41) Querques, F.; D'Agostino, A.; Cozzolino, C.; Cozzuto, L.; Lombardo, B.; Leggiero, E.; Ruosi, C.; Pastore, L. Identification of a Novel Transcription Factor Required for Osteogenic Differentiation of Mesenchymal Stem Cells. *Stem Cells Dev.* **2019**, *28*, 370−383.

(42) Cui, J.; Yin, Y.; Ma, Q.; Wang, G.; Olman, V.; Zhang, Y.; Chou, W.-C.; Hong, C. S.; Zhang, C.; Cao, S.; Mao, X.; Li, Y.; Qin, S.; Zhao, S.; Jiang, J.; Hastings, P.; Li, F.; Xu, Y. Comprehensive Characterization of the Genomic Alterations in Human Gastric Cancer. *Int. J. Cancer* **2015**, *137* (1), 86−95.

(43) Kanwal, M.; Ding, X.-J.; Ma, Z.-H.; Li, L.-W.; Wang, P.; Chen, Y.; Huang, Y.-C.; Cao, Y. Characterization of Germline Mutations in Familial Lung Cancer from the Chinese Population. *Gene* **2018**, *641*, 94−104.

(44) Liang, Y.; Jiang, L.; Zhong, X.; Hochwald, S. N.; Wang, Y.; Huang, L.; Nie, Q.; Huang, H.; Xu, J.-F. Discovery of Aberrant Alteration of Genome in Colorectal Cancer by Exome Sequencing. *Am. J. Med. Sci.* **2019**, *358* (5), 340−349.

(45) Liu, P.; Morrison, C.; Wang, L.; Xiong, D.; Vedell, P.; Cui, P.; Hua, X.; Ding, F.; Lu, Y.; James, M.; Ebben, J. D.; Xu, H.; Adjei, A. A.; Head, K.; Andrae, J. W.; Tschannen, M. R.; Jacob, H.; Pan, J.; Zhang, Q.; Van den Bergh, F.; Xiao, H.; Lo, K. C.; Patel, J.; Richmond, T.; Watt, M.-A.; Albert, T.; Selzer, R.; Anderson, M.; Wang, J.; Wang, Y.; Starnes, S.; Yang, P.; You, M. Identification of Somatic Mutations in Non-Small Cell Lung Carcinomas Using Whole-Exome Sequencing. *Carcinogenesis* **2012**, *33* (7), 1270−1276.

(46) Park, S.-M.; Seo, E.-H.; Bae, D.-H.; Kim, S. S.; Kim, J.; Lin, W.; Kim, K.-H.; Park, J. B.; Kim, Y. S.; Yin, J.; Kim, S.-Y. Phosphoserine Phosphatase Promotes Lung Cancer Progression through the Dephosphorylation of IRS-1 and a Noncanonical L-Serine-Independent Pathway. *Mol. Cells* **2019**, *42* (8), 604−616.

(47) Kutschera, S.; Weber, H.; Weick, A.; De Smet, F.; Genove, G.; Takemoto, M.; Prahst, C.; Riedel, M.; Mikelis, C.; Baulande, S.; Champseix, C.; Kummerer, P.; Conseiller, E.; Multon, M.-C.; Heroult, M.; Bicknell, R.; Carmeliet, P.; Betsholtz, C.; Augustin, H. G. Differential Endothelial Transcriptomics Identifies Semaphorin 3G as a Vascular Class 3 Semaphorin. *Arterioscler., Thromb., Vasc. Biol.* **2011**, *31* (1), 151−159.

(48) Yu, R.; Ma, L.; Li, J.; Gu, J.; Shi, Q.; Yu, R. Effects of SEMA3G on Migration and Invasion of Glioma Cells. *Oncol. Rep.* **2012**, *28* (1), 269−275.

(49) Ishibashi, R.; Takemoto, M.; Akimoto, Y.; Ishikawa, T.; He, P.; Maezawa, Y.; Sakamoto, K.; Tsurutani, Y.; Ide, S.; Ide, K.; Kawamura, H.; Kobayashi, K.; Tokuyama, H.; Tryggvason, K.; Betsholtz, C.; Yokote, K. A Novel Podocyte Gene, Semaphorin 3G, Protects Glomerular Podocyte from Lipopolysaccharide-Induced Inflammation. *Sci. Rep.* **2016**, *6* (1), 25955.

(50) Oleari, R.; André, V.; Lettieri, A.; Tahir, S.; Roth, L.; Paganoni, A.; Eberini, I.; Parravicini, C.; Scagliotti, V.; Cotellessa, L.; Bedogni, F.; De Martini, L.; Corridori, M. V.; Gulli, S.; Augustin, H. G.; Gaston-Massuet, C.; Hussain, K.; Cariboni, A. A Novel SEMA3G Mutation in Two Siblings Affected by Syndromic GnRH Deficiency. *Neuroendocrinology* **2021**, *111* (5), 421−441.

(51) Seo, H.; Lee, I.-S.; Park, J. E.; Park, S. G.; Lee, D. H.; Park, B. C.; Cho, S. Role of Protein Tyrosine Phosphatase Non-Receptor Type 7 in the Regulation of TNF-*α* Production in RAW 264.7 Macrophages. *PLoS One* **2013**, *8* (11), No. e78776.

(52) Wang, F.; Wang, X.; Liu, L.; Deng, S.; Ji, W.; Liu, Y.; Wang, X.; Wang, R.; Zhao, X.; Gao, E. Comprehensive Analysis of PTPN Gene Family Revealing PTPN7 as a Novel Biomarker for Immuno-Hot Tumors in Breast Cancer. *Front. Genet.* **2022**, *13*, 981603.

(53) Hansen, C. G.; Nichols, B. J. Exploring the Caves: Cavins, Caveolins and Caveolae. *Trends Cell Biol.* **2010**, *20* (4), 177−186.

(54) Lo, H. P.; Lim, Y.-W.; Xiong, Z.; Martel, N.; Ferguson, C.; Ariotti, N.; Giacomotto, J.; Rae, J.; Floetenmeyer, M.; Moradi, S. V.; Gao, Y.; Tillu, V. A.; Xia, D.; Wang, H.; Rahnama, S.; Nixon, S. J.; Bastiani, M.; Day, R. D.; Smith, K. A.; Palpant, N. J.; Johnston, W. A.; Alexandrov, K.; Collins, B. M.; Hall, T. E.; Parton, R. G. Cavin4

Interacts with Bin1 to Promote T-Tubule Formation and Stability in Developing Skeletal Muscle. *J. Cell Biol.* **2021**, *220* (12), No. e201905065.

(55) Gupta, R.; Toufaily, C.; Annabi, B. Caveolin and Cavin Family Members: Dual Roles in Cancer. *Biochimie* **2014**, *107*, 188−202.

(56) Felipe-Abrio, B.; Carnero, A. The Tumor Suppressor Roles of MYBBP1A, a Major Contributor to Metabolism Plasticity and Stemness. *Cancers* **2020**, *12* (1), 254.

(57) Doello, S.; Liang, Z.; Cho, I. K.; Kim, J. B.; Li, Q. X. Cytotoxic Effects of 24-Methylenecyloartanyl Ferulate on A549 Nonsmall Cell Lung Cancer Cells through MYBBP1A Up-Regulation and AKT and Aurora B Kinase Inhibition. *J. Agric. Food Chem.* **2018**, *66* (14), 3726−3733.

(58) Li, F.-N.; Zhang, Q.-Y.; Li, O.; Liu, S.-L.; Yang, Z.-Y.; Pan, L.-J.; Zhao, C.; Gong, W.; Shu, Y.-J.; Dong, P. ESRRA Promotes Gastric Cancer Development by Regulating the CDC25C/CDK1/CyclinB1 Pathway via DSN1. *Int. J. Biol. Sci.* **2021**, *17* (8), 1909−1924.

(59) Mukherjee, T. K.; Malik, P.; Hoidal, J. R. The Emerging Role of Estrogen Related Receptorα in Complications of Non-small Cell Lung Cancers (Review). *Oncol. Lett.* **2021**, *21* (4), 258.

(60) Liu, X.; Shi, D.; Zhou, S.; Liu, H.; Liu, H.; Yao, X. Molecular Dynamics Simulations and Novel Drug Discovery. *Expert Opin. Drug Discovery* **2018**, *13* (1), 23−37.

(61) Lobanov, M. Yu.; Bogatyreva, N. S.; Galzitskaya, O. V. Radius of Gyration as an Indicator of Protein Structure Compactness. *Mol. Biol.* **2008**, *42* (4), 623−628.

(62) Durham, E.; Dorr, B.; Woetzel, N.; Staritzbichler, R.; Meiler, J. Solvent Accessible Surface Area Approximations for Rapid and Accurate Protein Structure Prediction. *J. Mol. Model.* **2009**, *15* (9), 1093−1108.

(63) Yue, P.; Li, Z.; Moult, J. Loss of Protein Structure Stability as a Major Causative Factor in Monogenic Disease. *J. Mol. Biol.* **2005**, *353* (2), 459−473.

(64) Amadei, A.; Linssen, A. B.; Berendsen, H. J. Essential Dynamics of Proteins. *Proteins* **1993**, *17* (4), 412−425.

(65) Maisuradze, G. G.; Liwo, A.; Scheraga, H. A. Relation between Free Energy Landscapes of Proteins and Dynamics. *J. Chem. Theory Comput.* **2010**, *6* (2), 583−595.

(66) Gorelov, S. V.; Titov, A. I.; Tolicheva, O.; Konevega, A. L.; Shvetsov, A. V. DSSP in Gromacs: Tool for Defining Secondary Structures of Proteins in Trajectories. *bioRxiv* **2023**, bioRxiv:2023.08.21.554196.

(67) Pan, Z.; Wang, K.; Wang, X.; Jia, Z.; Yang, Y.; Duan, Y.; Huang, L.; Wu, Z.-X.; Zhang, J.-Y.; Ding, X. Cholesterol Promotes EGFR-TKIs Resistance in NSCLC by Inducing EGFR/Src/Erk/SP1 Signaling-Mediated ERRα Re-Expression. *Mol. Cancer* **2022**, *21* (1), 77.

(68) Xia, H.; Dufour, C. R.; Giguère, V. ERRα as a Bridge Between Transcription and Function: Role in Liver Metabolism and Disease. *Front. Endocrinol.* **2019**, *10*, 206.