EJMO

DOI: 10.14744/ejmo.2023.33189
EJMO 2023;7(4):334–344

Research Article

Identification of High-Risk Single Nucleotide Polymorphisms (SNPs) of Epidermal Growth Factor Receptor (EGFR) and Their Interaction with Various TKI Drugs

Ananya Sharma,[1] Raman Thakur,[2] Shikha Mittal,[1] Jata Shankar[1]

[1]Genomic Laboratory, Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Waknaghat, Solan, Himachal Pradesh, India
[2]Department of Medical Laboratory Science, Lovely Professional University, Jalandhar, Punjab, India

Abstract

Objectives: Epidermal growth factor receptor (EGFR) is the membrane receptor of tyrosine kinase family that plays crucial role in cell growth, cell division and cell survival. Upregulation of EGFR gene has been seen in various cancer types due to mutations. As non-synonymous single nucleotide polymorphisms (nsSNPs) are responsible for half of the genetic variations that are responsible of human diseases, it is crucial to analyze putative functional nsSNPs. Therefore, we aimed to identify nsSNPs of EGFR gene and evaluate their effect on its protein receptor. We also docked the mutant type protein structures with common tyrosine kinase inhibitor (TKIs) to showcase their stability with one another.

Methods: We observed five novel nsSNPs (E330K, K745R, R962H, R675Q and S752Y) that are present in different domains of EGFR using various bioinformatics tools and simultaneously predicted their deleterious effects on EGFR. Furthermore, docking studies were carried out with three of the common TKIs (erlotinib, gefitinib, canertinib).

Results: Mutant type K745R was predicted to be potentially more damaging than other mutants due to its presence in highly conserved region of EGFR protein receptor and its ability to affect protein stability.

Conclusion: As this study is the first comprehensive study of these novel nsSNPs of EGFR, the results of this study would be crucial for future studies, drug discovery and development of personalized medicine. Although along with in-silico characterisation of nsSNPs clinical population-based studies are essential.

Keywords: Epidermal growth factor receptor (EGFR), Single Nucleotide Polymorphism (SNP), Tyrosine Kinase Inhibitors (TKIs), Non-Small Cell Lung Cancer (NSCLCs)

Cite This Article: *Sharma A, Thakur R, Mittal S, Shankar J. Identification of High-Risk Single Nucleotide Polymorphisms (SNPs) of Epidermal Growth Factor Receptor (EGFR) and Their Interaction with Various TKI Drugs. EJMO 2023;7(4):334–344.*

We are aware about the journey of normal cells to hyperplasia and dysplasia which leads to malignancies and ultimately cancer which spreads to other body parts by the process of metastasis. In 2022, 1,918,030 new cancer cases and 609,360 deaths due to cancer has been projected in United States. According to American cancer society, approximately 350 deaths occur per day from lung cancer which makes it the leading cause of death due to cancer worldwide. Whereas, prostate cancer in males and breast cancer in females are the estimated new cancer cases with leading score of 268,490 (27%) and 287,850 (31%) respectively.[1] In recent times as researchers are taking pains to develop an effective treatment protocol. Identification and characterisation of various tumor sup-

Address for correspondence: Raman Thakur & Jata Shankar. Genomic Laboratory, Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Waknaghat, Solan, Himachal Pradesh, India
Phone: +91 1792 239225 **E-mail:** ramanthakurnegwal@gmail.com, jata.shankar@juit.ac.in

Submitted Date: May 06, 2023 **Revision Date:** July 24, 2023 **Accepted Date:** July 30, 2023 **Available Online Date:** December 29, 2023

pressor genes is crucial as loss of their activity for tumor suppression is the benchmark of cancer.[2] This study focuses mainly on epidermal growth factor receptor (EGFR), located at the short arm of the chromosome 7 (7p11. 2). The accidental discovery of this protein receptor was done in 1959 by Stanley Cohen and Rita Levi-Montalcini when they were examining the effects of nerve growth factor (NGF) on new-born mice and noticed the unexpected precocious effects, they try to focus on the sole component responsible for those side-effects, they isolated and termed it as epidermal growth factor (EGF), a 6000 Da polypeptide, as it enhances the epidermal growth in vivo as well as in vitro in animal cells. They showed this growth factor binds to a receptor called epidermal growth factor receptor (EGFR).

EGFR also known as ErbB1/ HER1 (human EGFR related), is from the tyrosine kinases class of enzymes that play role in epithelial cell physiology, promotes various pro-oncogenic processes which includes angiogenesis, metastasis, adhesion, cell proliferation, cell motility and inhibition of apoptosis, which are the main reasons for cancer progression upon mutation and thus also known as proto-oncogene. [3] It is a single-chain transmembrane receptor tyrosine kinase having binding site for epidermal growth factor (EGF), transforming growth factor-α (TGF-α), epiregulin and four other ligands on its extracellular domain other than that it has single pass transmembrane domain and an intracellular domain as well, total protein encoding for about 1210 amino acids.[4] Upon ligand binding the receptor undergoes dimerization forming homo- or heterodimers. This receptor and its ligands have been seen upregulated in various human cancer. Normal cells express 40,000-100,000 EGFR on them whereas cancerous cells express 10^6 EGFR per cell. [5] Mutations in EGFR shows the therapeutic importance as it predicts the efficacy of EGFR inhibitors strongly with the response rate of more than 70%.[6]

The upregulation of this receptor is seen in various cancer type due to mutations found at specific mutational hotspots in its various regions/domains.[7] The mutation can be single nucleotide polymorphisms (SNPs) which are the changes only in a single nucleotide that occur approximately after every 1000 nucleotides in a single population. In total more than 30,000 SNPs have been recorded till now of EGFR that may have a role in cancer progression. SNPs that alter the encoded amino acids are known as non-synonymous (nsSNPs), they lead to the change structure of the protein which may alter its function. These nsSNPs are responsible for half of all the genetic alterations related to human diseases.[8-10]

In this study we have used public datasets and bioinformatics tools to predict the deleterious nsSNPs of EGFR gene and to understand consequent protein effect on its structure and function. Apart from that we have also dock the variants with three of the common tyrosine kinase inhibitors (TKIs) which are used in cancer therapy and analysed their binding energies. This study would be essential for future studies in this regard.

Methods

SNP Data Mining for EGFR Human Gene

The SNP data file for the human EGFR gene was obtained from the NCBI-hosted dbSNP database (https://www.ncbi.nlm.nih.gov/snp/), a free repository of single nucleotide genetic variants for various species. The protein sequence (FASTA format) for the wild type EGFR gene was obtained using Uniprot database (https://www.uniprot.org/uniprotkb/P00533/) entry(accession no- P00533).

Prediction of Deleterious SNPs

Among these SNPs from the dataset, nsSNPs were identified. Therefore, several in-silico approaches were employed to determine if SNPs had a detrimental influence on protein structure and function.

Sorting Intolerant from Tolerant (SIFT) (https://sift.bii.astar.edu.sg/www/SIFT dbSNP.html) predicts deleterious nsSNPs based on sequence homology and the physical characteristics of amino acids, this analyser predicts if SNPs are deleterious or non-deleterious. The SIFT score below 0.05 implies that the SNP has detrimental impact on protein structure.[11]

Polyphen-2 (Polymorphism Phenotyping v2) (http://genetics.bwh.harvard/pph2/)is an automated algorithm that predicts the effect of amino acid substitution on protein structure. It forecasts the outcomes using a scale ranging from 0.00 to 1.00, where > 0.15 suggests "possibly damaging," > 0.85 indicates "probably destructive," and remaining mutations imply "benign."[12]

PanthercSNP 17.0 (Protein analysis through evolutionary relationship-coding SNP v17) tool (http://www.pantherdb.org/tools/csnpScore.do) is based on evolutionary relationship, molecular functions and interaction with other proteins. This web-based tool provides PSEP (position-specific evolutionary preservation) which evaluate how long a position in a particular protein has been preserved. As longer the position has been preserved more damage it would cause. This tool converts PSEP to probability of deleterious effect (Pdel) scores.[13]

SNAP2 (Screening for Nonacceptable Polymorphisms) (https://rostlab.org/services/snap/) differentiates between

neutral and effect variants by providing the scores at the end of the prediction which varies from -100 (strong neutral prediction) to +100 (strong effect prediction). Stronger the effect score for a variant more deleterious effect it could show.[14]

Apart from that Predict SNP (https://loschmidt.chemi.muni.cz/predictsnp/) is a classifier for prediction of deleterious SNPs and predict a collective data from 8 different in silico SNP predicting tools.[15]

Prediction of Protein Stability

A web-based server I-Mutant suite (version-3) (http://gpcr2.Biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi) was used to predict the protein stability. This is a support vector machine (SVM) based in silico prediction tool for predicting protein stability after a single point mutation. The initial need for conducting the predictions is the protein sequence, and the output is the free energy change (DDG) value, which is the difference between the free energy of the mutant protein and the wild type protein. A DDG value below 0 implies that the mutant protein is less stable than its wild-type counterpart, while a DDG value above 0 suggests that the altered protein is more stable. This tool also predicts the sign of Gibbs free energy drop or increase using the reliability index (RI) based on amino acid change between 0 and 10, where RI-0 denotes the lowest dependability and RI-10 the most.All five nsSNPs found out to be less stable than the wild type EGFR.Throughout the method, the temperature and pH were set at 24°C and 7 for all SNPs, respectively.[16]

Evaluation of Conservation Profile by ConSurf

Consurf a bioinformatics tool was used to know the evolutionary conservation of nsSNPs, this tool calculates the data based on phylogenetic relations between homologous sequences. This tool uses colour scheme to represent the conservation score that varies from 1 to 9 with labels variable, average and highly conserved for each amino acid position.[17]

Protein-Protein Interaction Prediction of EGFR

PPIs plays very crucial role in cellular functions in every organism therefore, understanding their interactions can lead to the development of better treatment protocols and optimization.

STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) (https://string-db.org/)

Database predict interactions include physical and functional associations and used to predict EGFR interaction with other proteins.[18]

Prediction of Domains of EGFR Protein and nsSNPs Positions

Two web-based tools, NCBI conserved domain search tool (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) and InterPro (https://www.ebi.ac.uk /interpro/) were used to figure out the domains of EGFR where the selected nsSNPs resides. These web-based servers require protein sequence in FASTA format as input and provide the amino acid range for the domains. The FASTA sequence for EGFR genewas retrieved from Uniprot (https://www.uniprot.org/uniprotkb /P00533/entry) (accession no- P00533). The receptor has three regions in total, the extracellular region, the transmembrane region and the intracellular region. The extracellular region is the ligand binding region comprising of further four domains. Domains I and III are leucine rich and play a role in ligand binding. Whereas, domain II and IV are cysteine rich and form homo- or heterodimers with other members of the family and form disulphide bonds to domain II, respectively. The transmembrane region is a small single pass hydrophobic domain seems to play a role in dimerization and helps the receptor to get anchored in the membrane. The intracellular region includes a flexible juxtamembrane segment, the tyrosine kinase domain and a C-terminal tail. The tyrosine kinase domains consist of two lobes, the N- (which is mainly beta sheets) and the C-lobes (mainly alpha helices) and an ATP binding sites between the two lobes, this plays role during transautophosphorylation.[19]

Prediction of 3-D Protein Structure of High Risk nsSNPs

For the 3-dimentional structure prediction of EGFR mutants Phyre 2.0 (http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index) in-silico tool was used.It is the free web-based service for prediction of protein structure using the principles of homology modelling. This web server uses Hidden Markov models (HMMs) for detecting and aligning sequences and provide with a 3-D structure. It requires the protein sequence in FASTA format which was downloaded from Uniprot for wilt type EGFR and alterations were made manually at the position of nsSNPs.The job was run in intensive mode to get the complete structure.[20] The structures later downloaded in PDB format and followed by energy minimisation in Yasara energy minimisation tool to improve the quality of predicted model.[21] Procheck (https://saves.mbi.ucla.edu/) web server is a well-known tool to describe the stereo-chemical quality of the protein structure, using that we obtained Ramachandran plot to observe the allowed and disallowed regions for amino acids of the predicted structures.[22]

Molecular Docking Studies

Energy minimised mutants and the wild type protein structure all were used for molecular docking studies. Three of the common tyrosine kinase inhibitors (TKIs) were selected and their compound information and structures were downloaded from PubChem (https://pubchem.ncbi.nlm.nih.gov/) database. Docking was performed using Autodock, autodock vina and MGL tools 1.5.6 and binding energies were compared. While preparing the protein structures wild type or mutant, kollman charges were incorporated, polar hydrogens were added and all the atoms were assigned to AD4 type. All the water molecules were removed and the prepared protein was saved in PDBQT format. Followed by ligand preparation, structures in PDB format of TKIs were converted to PDBQT format using openBabel application as this format is functional in Autodock v.1.2. After preparation protein-ligand complex was saved in PDBQT format. The grid dimensions varied for different complexes of ligand and the mutants. The results were binding free energy which was obtained from autodock vina commands. The ligand with the lowest binding energy visualised with the protein structures using discovery studio visualizer (DSV).

Results

SNP Data File Retrieval

SNPs for human EGFR gene was retrieved from DbSNP dataset (https://www.ncbi.nlm.nih.gov/snp) shows 80435 SNPs out of which2976 were missense (nsSNPs), 2117 were non-coding transcripts, 1218 were synonymous, 74094 were intronic, 5 were initiator codon variants, 80 were inframe insertions, 70 were inframe indels and 24 were inframe deletions. In this study only nsSNPs were considered. 77 SNPs were filtered out from 2976 missense mutations based on their clinical significance. Number of SNP data is present in Table1 and their distribution is given in Figure 1.

Table 1. SNPs filtered out using various in-house filters

Clinical Significance Criteria	No. of SNPs
Benign	24
Benign, likely-benign	1
Likely-benign	29
Likely-benign, benign	3
Likely pathogenic	17
Pathogenic	3
Total	77

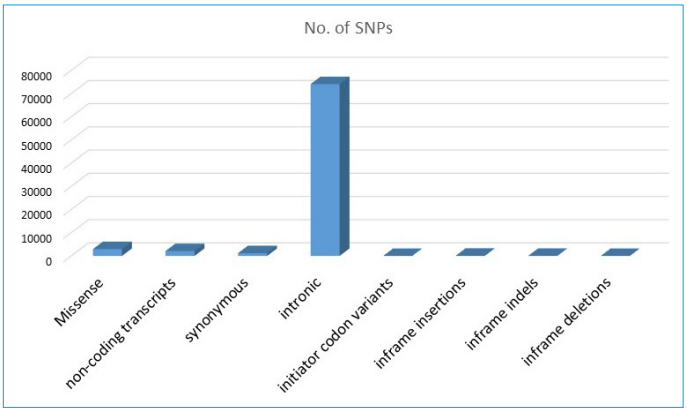

Figure 1. Represents the distribution of SNPs of human EGFR gene in different functional classes retrieved from dbSNP database.

Analysis of nsSNPs to Predict Deleterious SNPs and Their Consequences

Those 77 SNPs were subjected to SIFT analysis and out of them 9 were found to be deleterious as having score less than 0.05. To cross validate the prediction of SIFT tool score, PolyPhen 2.0, Panther17.0 and SNAP2 were further used. Predict SNP tool was used to get the collective results from 8 other SNP tools (Table 2 and 3). On performing the comparative analyses of all the tools, variant K745R is predicted deleterious by each of the tool.

Analysis of Structural Stability of Protein

I-mutant suite a web-based server, was used to predict the effect of mutants on protein stability as compared with the wild type protein. Out of 5 mutants 4 (E330K, K745R, R962H, R675Q) were found to decrease the stability of the protein whereas mutant S752Y seems to increase the protein stability (Table 4).

Conservation Profile of Deleterious nsSNPs in EGFR

Using ConSurfweb-based server, all 5 nsSNPs were analysed to predict evolutionary conservation and to know whether the mutant is exposed or buried and functional or structural. Based on conservation scores out of 5 nsSNPs, 2 were found to be highly conserved residues (K745R, S752Y), 2 predicted to be moderately conserved (R962H, R675Q) and 1 considered as variable (E330K). Furthermore, all the variants were predicted as exposed and functional (Table 4). On comparison of I-mutant suite and ConSurf data we estimated the out of 5 nsSNPs 2 nsSNPs (K745R and R962H) are potentially be highly damaging.

Protein-Protein Interaction Analysis

Using STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) database various interactions of EGFR were revealed. EGFR interacts with RASA1 (Ras GTPase-

Table 2. Deleterious SNPs that were selected after SNP datamining

SNP ID	Alleles	Amino acid change	SNP Position (nt)	SNP Position (AA)	Region	PolyPhen-2 score	SIFT score	Panther 17.0	SNAP2 score
rs121913433	A>G	K>R	55174771	745	CDS	1.000	0.002	Probably damaging	61
rs121913464	C>A	S>Y	55174792	752	CDS	0.992	0.01	Probably damaging	48
rs139429793	G>A	E>K	55155928	330	CDS	1.000	0.02	Probably damaging	-24
rs144496976	G>A	R>H	55200352	962	CDS	1.000	0.04	Probably damaging	-16
rs150423237	G>A	R>Q	55173087	675	CDS	1.000	0.009	Probably damaging	43

Table 3. Prediction of deleterious nsSNPs using Predict SNP

Mutation	Predict SNP		MAPP		PhD-SNP	
	Deleterious/ Neutral	Percentage	Deleterious/ Neutral	Percentage	Deleterious/ Neutral	Percentage
E330K	Neutral	75%	-	-	Neutral	58%
K745R	Deleterious	76%	Deleterious	51%	Deleterious	81%
R675Q	Deleterious		-	-	Neutral	58%
S752Y	Deleterious	72%	Neutral	65%	Deleterious	77%
R962H	Neutral	60%	Neutral	76%	Neutral	78%

Table 4. Analysis of protein stability and evolutionary conservation profile of nsSNPs using I-mutant suite and ConSurf web servers

nsSNPs	AA position	I-mutant suite			ConSurf		
		DDG (kcal/mol)	DDG sign	RI	Conservation score	Buried/ Exposed	Functional/ Structural
rs121913433	K745R	-0.29	Decreased	5	9	Exposed	Functional
rs121913464	S752Y	-0.01	Increased	4	9	Exposed	Functional
rs144496976	R962H	-1.49	Decreased	9	8	Exposed	Functional
rs139429793	E330K	-0.51	Decreased	8	4	Exposed	-
rs150423237	R675Q	-1.08	Decreased	9	7	Exposed	-

activating protein 1), HBEGF (Proheparin-binding EGF-like growth factor), TGFA (Protransforming growth factor alpha), PLCG1 (1-phosphotidylinositol 4,5-bisphosphate phosphodiesterase gamma-1), EGF (pro-epidermal growth factor), EREG (Epiregulin), CBL (E3 ubiquitin-protein ligase CBL), HSP90AA1 (Heat shock protein HSP 90-alpha), STAT3 (Signal transducer and activator of transcription 3), CDH1 (Cadherin-1) Figure 2.

Domain Prediction of nsSNPs of EGFR

To identify the domains where the mutant nsSNPs lies in EGFR gene, NCBI conserved domain search tool and InterPro was used (Table 5 and 6). These platforms accept protein sequence in FASTA format or protein ID as input to extrapolate the domains and motifs of particular protein.

3D Mutant Protein Structure Prediction and Energy Minimisation

Three-dimensional structure of the mutant proteins were predicted in PDB format using Phyre 2.0 web server which require amino acid sequence of EGFR as input which was retrieved from Uniprot. Energy minimisation of predicted structures was performed using Yasara web-based tool, Z-score and minimised energy values are summarised in table 7. The stereochemical quality of variant protein structures and Ramachandran plot was predicted using PROCHECK.

Table 5. Domains of EGFR and their AA location

Receptor_L_Domain	Furin-like	Receptor_L_Domain	GF_recep_IV	TM_ErbB1	PTKc_EGFR
57-168 AA	185-335 AA	361-481 AA	505-637 AA	634-677 AA	704-1016 AA

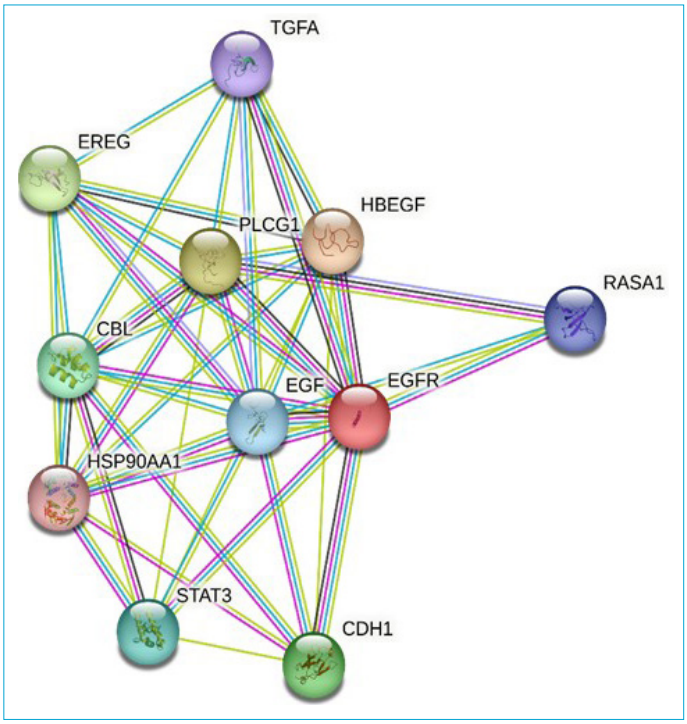

Figure 2. Represents the Protein-Protein interaction network of human EGFR using STRING server.

Table 6. Predicted domains of nsSNPs

S No.	nsSNPs	AA position	Domain
1	rs121913433	K745R	Tyrosine kinase domain
2	rs121913464	S752Y	Tyrosine kinase domain
3	rs139429793	E330K	Domain-2 of extracellular region
4	rs144496976	R962H	Tyrosine kinase domain
5	rs150423237	R675Q	Transmembrane domain

Table 7. Energy minimisation values from Yasara

Mutants	Start energy (KJ/mol)	Start Z-score	End energy (KJ/mol)	End Z-score
E330K	34029401451.9	-2.87	-579651.7	-1.86
K745R	364830555092.8	-3.04	-579487.1	-1.88
R675Q	5670043650501.7	-2.96	-581581.6	-1.81
S752Y	1085669381890.6	-2.91	-576903.9	-1.78
R962H	2549405494.5	-2.81	-580408.8	-1.78

Molecular Docking Results

All the 5 mutants and the wild type protein structure were dock with the three selected TKIs which are summarised in table 9 and their 2-D structure in figure 3 retrieved from PubChem database https://pubchem.ncbi.nlm.nih.gov/. Protein and ligand were prepared prior to the docking and binding energies of 9 ligand conformations were retrieved by performing Autodock vina v.1.2. Lowest binding energies were considered for further visualisation in discovery studio visualizer as lower the energy higher the stability. This will give us an insight of the stable conformations of common TKIs for novel nsSNPs which is useful provides useful data for selection of targets in future for the treatment of cancer. Consider table 10. While visualization in DSV all the domains of the EGFR protein were labelled using different colour to avoid confusion. Of extracellular domain, domain 1 is labelled with red, domain 2 labelled with yellow, domain 3 labelled with blue, domain 4 is labelled with green, transmembrane domain labelled with pink and tyrosine kinase domain is labelled with purple colour with schematic style and the ligand labelled with dark pink colour in CPK style, Figure 4a-4f. Apart from that interacting amino acids of ligand and protein were also known using DSV, Figure 5a,b.

Discussion

EGFR is the receptor protein that present on the surface of almost every cell in the human body and bind with the epidermal growth factor, tumour necrosis factor-α and several other ligands. This protein receptor involves in various signalling pathways that control cell division as well as cell survival. The mutation in the EGFR gene led to higher expression of EGFR protein receptor on some cells and therefore help those cells to divide rapidly.[5] Hence its overexpression is found to be associated with lung cancer, breast cancer, glioblastomas, metastatic colorectal cancer, head and neck cancer, prostate and ovarian cancer and other cancer types.[6, 23, 24] In European patients of lung adenocarcinoma, 5-10% of mutations in EGFR gene has been observed.[25] Mutations in EGFR seems to cause drug resistance and relapse of dis-

Table 8. Percentage distribution of variant for various regions of Ramachandran plot using PROCHECK

Mutants	% of residues in favoured regions	% of residues in additionally allowed regions	% of residues in generously allowed regions	% of residues in disallowed regions
E330K	84.5%	14%	0.8%	0.7%
K745R	84.7%	13.2%	1.3%	0.8%
R675Q	84.2%	13.8%	1.3%	0.7%
S752Y	85.3%	12%	2%	0.9%
R962H	85.3%	12.8%	1%	1%

Table 9. Various compound properties of the selected TKIs

Compound	Molecular weight (g/mol)	H-bond donor	H-bond acceptor	Log P	Pubchem CId	Molecular formula
Erlotinib	393.4	1	7	2.7	176870	$C_{22}H_{23}N_3O_4$
Gefitinib	446.9	1	8	4.1	123631	$C_{22}H_{24}ClN_4O_3$
Canertinib	485.9	2	8	3.9	156414	$C_{24}H_{25}ClFN_5O_3$

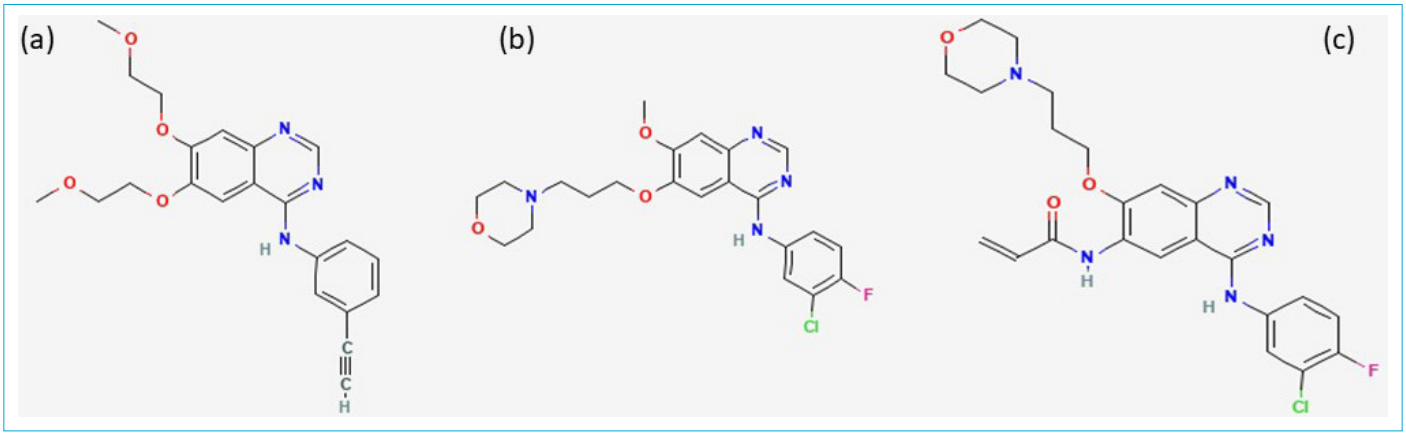

Figure 3. Represent the structure of some tyrosine kinase inhibitors used for docking studies.

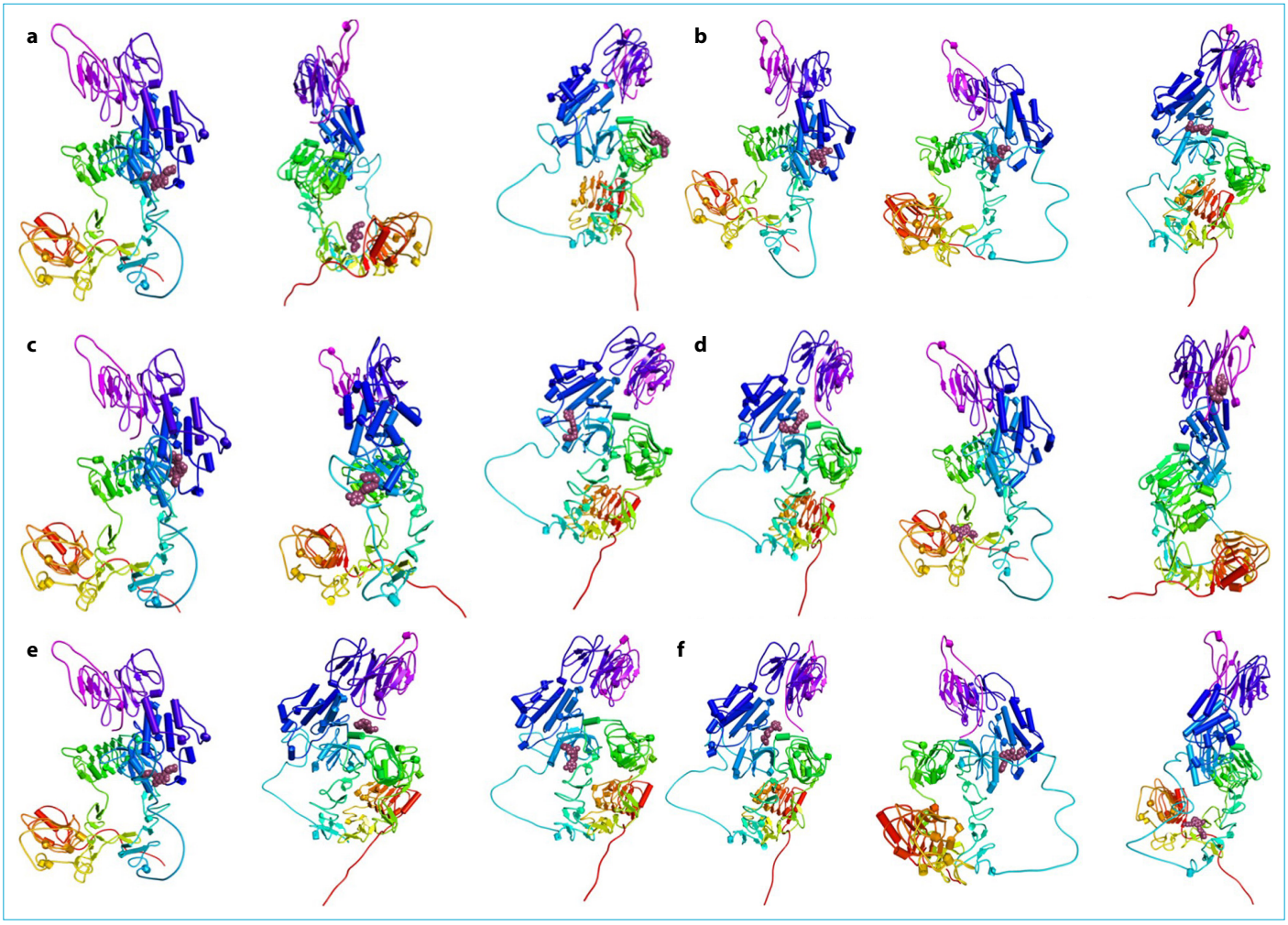

Figure 4 (a-f). Represent the interaction of wild type and mutant type protein structures with various tyrosine kinase inhibitors.

Table 10. Values of kollman charges and various binding energies of ligands

Protein	Lowest binding energy with erlotinib	Lowest binding energy with gefitinib	Lowest binding energy with canertinib	Kollman charges
Wild type	-7.5	-8.8	-7.6	34.0
E330K	-6.5	-7.0	-7.8	20.322
K745R	-6.4	-7.8	-7.7	18.322
R675Q	-6.4	-7.3	-7.4	16.983
S752Y	-6.5	-7.2	-7.8	18.322
R962H	-6.7	-7.4	-7.1	18.322

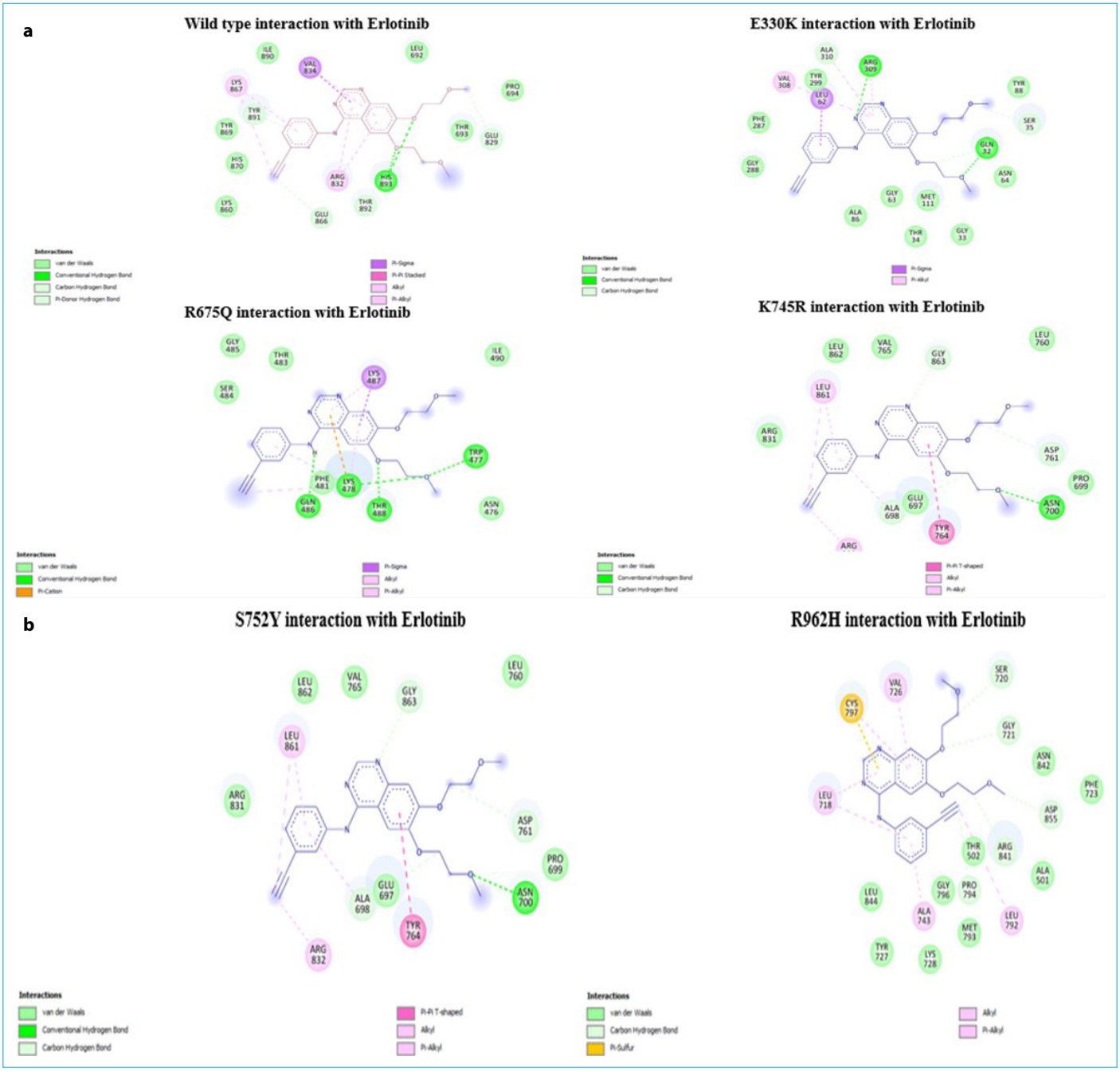

Figure 5 (a, b). Represent the interacting amino acids between wild and mutant type proteins and ligands (tyrosine kinase inhibitors).

ease in patients with non-small cell lung cancer.[26] In recent years the effect of various deleterious SNPs has been demonstrated in different diseases and disorders including cancer. Further in-silico studies havebeen carried out to identify new deleterious SNPs among different genes that may contributes for various human diseases and can be tested clinically.[27] Despite the impact of EFGR has been investigated in association with various cancer types but in-silico analysis of deleterious SNPs in EGFR gene and its impact on protein structure and function is remains uncharacterized. Therefore, to enhance the effectiveness of detecting more deleterious SNPs we combined several computer-based approaches and predict the effect of deleterious SNPs on EGFR. Hence, in this study we predicted the deleterious SNPs using different prediction tools such as, SIFT, PolyPhen 2.0, Panther, SNAP2 and Predict SNP. Using these algorithms, we selected 5 nsSNPs as high risk SNPS that were predicted as most deleterious by almost all used algorithms and further subjected to different analysis. This prediction was done on the bases of prediction scores provided by these in-silico algorithms. While comparing the scores for different in silico tools, mutant K745R predicted to be damaging. Further I-mutant suite, a web-based tool was used to predict the stability of selected nsSNPs which is required for structural and functional activity of proteins. Four out of 5 mutants (E330K, K745R, R962H, R675Q) were found to be less stable than that of wild type structure, which suggested that these mutants may interrupt EGFR protein function, whereas mutant S752Y observed to enhance the protein stability. These alteration in the protein structure may cause misfolding and degradation of proteins.[28] To determine the evolutionary conservation among nsSNPs ConSurf were performed. Again, on comparative analysis of scores obtained from ConSurf and I-mutant suite, K745R variant was most damaging with conservation score of 9, and decrease in protein stability. Other variant, R962H, also predicted as damaging with conservation score of 8. To understand the impact of point mutation on the cellular process of particular gene product, protein-protein interaction network is crucial. Using STRING database protein-protein interaction of EGFR gene was illustrated and revealed that disruption in any of these pathways might leads to diseases conditions. EGFR primarily interacts with ligands such as EGF, TGFA, EREG, HBEGF to initiate the signaling response.[4] Other than these, EGFR also interact with RASA1, PLCG1, CBL, HSP90AA1, STAT3 and CDH1. Overexpression of any of these proteins may lead to tumor progression.

Location of these nsSNPs in conserved domain revealed mutant K745R, S752Y and R962H belonged to tyrosine kinase domain, whereas mutant E330K belonged to domain-2 of extracellular region and mutant R675Q to transmembrane domain of EGFR gene. As all receptor tyrosine kinases, EGFR also comprise of large extracellular region, a single-spanning transmembrane domain, an intracellular region which consists of intracellular juxtamembrane region, a tyrosine kinase domain and a C-terminal regulatory region. The extracellular region comprises of four domains: I-IV, domain I and III are leucine rich and take part in ligand binding and domain II and IV are cysteine rich and participates in dimer formation.[29] On ligand binding there is a conformational change in the receptor followed by its dimerization and activation of tyrosine kinase domain followed transphosphorylation of tyrosine residues.[30] There are more than 12 tyrosine residues in cytoplasmic domain of EGFR which phosphorylates, and provide binding site for various proteins that are linked to various signal transduction pathways.[31] All of these five nsSNPs in this study are present in these regions which are important for functioning of the receptor. Tyrosine kinase domain responsible of transphosphorylation, domain-2 of extracellular region and transmembrane region takes part in dimerization. Mutation in various domains may affect the functioning of the EGFR, and needs their screening in cancer patients. Analysis on predicted protein structure containing these mutants showed change in energy. Allowed and disallowed regions of Ramachandran plot for amino acid residues was illustrated using the help of PROCHECK which predicted almost 85% of all the mutant types amino acids residues lies in favourable region. After energy minimisation of all the predicted protein structures using Yasara tool, all the consequent structures were dock on three common TKIs (erlotinib, gefitinib and canertinib). Erlotinib and Gefitinb are the first generation, reversible TKIs that are approved for the patients with NSCLCs.[25] Erlotinib is revealed to delay the cancer progression, improve the quality of life and increase the survival rates as first-line treatment compared to standard chemotherapy in patients with classic mutations of exon 21 (L858R) and exon 19 substitution and microdeletion respectively.[32] Canertinib is an irreversible inhibitor of RTKs including EGFR, binds to the ATP-binding site.[33] Clinical trials of canertinibhave been conducted with various other tumor types such as NSCLCs[34] and breast cancer.[35] However, it shows modest effects in clinical studies of ovarian cancer when used alone.[36] While dock with erlotinib and gefitinib drugs wild type structure has the lowest binding energies while R675Q and K745R mutants showed highest binding energies and E330K mutant shows the lowest binding energy and R962H mutant shows the highest binding energy when dock with canertinib. Mutant K745R (rs121913433) of tyrosine kinase domain of EGFR predicted to be most deleterious after overall analysis and found to be involved in squamous cell carcinoma of the head and

neck.[37] This mutation present in ATP binding site of tyrosine kinase domain of EGFR and may confer resistance to TKIs due to stabilisation of residues that participates in binding of ATP and TKIs.[38] However, more robust investigation is required to understand this variant closely and to identify its association with other cancer types.

Conclusion

As a proto-oncogene, EGFR involved in basic cellular functions such as cell growth and proliferation, alteration in the same leads to various cancers. This is the first comparative study of these 5 nsSNPs of EGFR (E330K, K745R, S752Y, R962H and R675Q). Among them we reported K745R as potentially damaging due to its presence in highly conserved region and ability of affect protein stability. It is found to be involved in Squamous cell carcinoma of the head and neck. However, in-silico investigation on large scale along with clinical trials is required to understand the effects of this nsSNPs more closely.

Disclosures

Ethics Committee Approval: No ethical approval required as no human and animal subject used in current study.

Peer-review: Externally peer-reviewed.

Conflict of Interest: None declared.

Authorship Contributions: Concept – R.T., J.S.; Design – R.T., J.S.; Supervision – J.S.; Materials – J.S.; Data collection &/or processing – A.S., S.M., R.T.; Analysis and/or interpretation – A.S., S.M., R.T.; Literature search – A.S., S.M., R.T.; Writing – A.S.; Critical review – R.T., J.S.

References

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. CA Cancer J Clin 2022;72:7–33.
2. Guo XE, Ngo B, Modrek AS, Lee WH. Targeting tumor suppressor networks for cancer therapeutics. Curr Drug Targets 2014;15:2–16.
3. Sigismund S, Avanzato D, Lanzetti L. Emerging functions of the EGFR in cancer. Mol Oncol 2018;12:3–20.
4. Singh B, Carpenter G, Coffey RJ. EGF receptor ligands: recent advances. F1000Research 2016;5:F1000.
5. Wee P, Wang Z. Epidermal growth factor receptor cell proliferation signaling pathways. Cancers Basel 2017;9:52.
6. Malik PS, Raina V. Lung cancer: Prevalent trends & emerging concepts. Indian J Med Res 2015;141:5–7.
7. Li AR, Chitale D, Riely GJ, Pao W, Miller VA, Zakowski MF, et al. EGFR mutations in lung adenocarcinomas: Clinical testing experience and relationship to EGFR gene copy number and immunohistochemical expression. J Mol Diagn 2008;10:242–8.
8. Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N, et al. Characterization of single-nucleotide polymorphisms in coding regions of human genes. Nat Genet 1999;22:231–8.
9. Thakur R, Shankar J. In-silico analysis revealed high-risk single nucleotide polymorphisms in human pentraxin-3 gene and their impact on innate immune response against microbial pathogens. Front Microbiol 2016;7:192.
10. Thakur R, Shankar J. Comprehensive in-silico analysis of high-risk non-synonymous snps in dectin-1 gene of human and their impact on protein structure. Curr Pharmacogenomics Pers Med 2017;15:144–55.
11. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res 2003;31:3812–4.
12. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods 2010;7:248–9.
13. Tang H, Thomas PD. PANTHER-PSEP: Predicting disease-causing genetic variants using position-specific evolutionary preservation. Bioinformatics 2016;32:2230–2.
14. Bromberg Y, Yachdav G, Rost B. SNAP predicts effect of mutations on protein function. Bioinformatics 2008;24:2397–8.
15. Bendl J, Stourac J, Salanda O, Pavelka A, Wieben ED, Zendulka J, et al. PredictSNP: Robust and accurate consensus classifier for prediction of disease-related mutations. PLoS Comput Biol 2014;10:e1003440.
16. 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 issue):W306-10.
17. Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic acids research. 2010;38(Web Server issue):529–33.
18. Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, et al. STRING 8 - A global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Res 2009;37(Database issue):412–6.
19. Jorissen RN, Walker F, Pouliot N, Garrett TP, Ward CW, Burgess AW. Epidermal growth factor receptor: Mechanisms of activation and signalling. Exp Cell Res 2003;284:31–53.
20. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. 2015;10:845–58.
21. Land H, Humble MS. YASARA: A tool to obtain structural guidance in biocatalytic investigations. Methods Mol Biol 2018;1685:43–67.
22. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: A program to check the stereochemical quality of protein structures. J Appl Crystallogr 1993;26:283–91.
23. Sasaki T, Hiroki K, Yamashita Y. The role of epidermal growth factor receptor in cancer metastasis and microenvironment. BioMed Res Int 2013;2013:546318.
24. Marzouq M, Nairouz A, Ben Khalaf N, Bourguiba-Hachemi S, Quaddorah R, Ashoor D, et al. Genetic variants of the EGFR li-

gand-binding domain and their association with structural alterations in Arab cancer patients. BMC Res Notes 2021;14:146.
25. Markóczy Z, Sárosi V, Kudaba I, Gálffy G, Turay ÜY, Demirkazik A, et al. Erlotinib as single agent first line treatment in locally advanced or metastatic activating EGFR mutation-positive lung adenocarcinoma (CEETAC): An open-label, non-randomized, multicenter, phase IV clinical trial. BMC Cancer 2018;18:598.
26. Zhang Y, Li Y, Wang Q, Su B, Xu H, Sun Y, et al. Role of RASA1 in cancer: A review and update. Oncol Rep 2020;44:2386–96.
27. Rozario LT, Sharker T, Nila TA. In-silico analysis of deleterious SNPs of human MTUS1 gene and their impacts on subsequent protein structure and function. PloS One 2021;16:e0252932.
28. Hossain MS, Roy AS, Islam MS. In-silico analysis predicting effects of deleterious SNPs of human RASSF5 gene on its structure and functions. Sci Rep 2020;10:14542.
29. Ferguson KM. Structure-based view of epidermal growth factor receptor regulation. Ann Rev Biophys 2008;37:353–73.
30. Iwata KK, Beard SE, Haley JD. Epidermal growth factor receptor (EGFR) inhibitor for oncology: Discovery and development of erlotinib. In Metcalf BW, Dillon S, editors. Target Validation in Drug Discovery. Burlington: Academic Press; 2007. p. 155–78.
31. Thelemann A, Petti F, Griffin G, Iwata K, Hunt T, Settinari T, et al. Phosphotyrosine signaling networks in epidermal growth factor receptor overexpressing squamous carcinoma cells. Mol Cell Proteomics 2005;4:356–76.
32. Zhou C, Wu YL, Chen G, Feng J, Liu XQ, Wang C, et al. Erlotinib versus chemotherapy as first-line treatment for patients with advanced EGFR mutation-positive non-small-cell lung cancer (OPTIMAL, CTONG-0802): A multicentre, open-label, randomised, phase 3 study. Lancet Oncol 2011;12:735–42.
33. Hassan W, Chitcholtan K, Sykes P, Garrill A. A combination of two receptor tyrosine kinase inhibitors, canertinib and PHA665752 compromises ovarian cancer cell growth in 3D cell models. Oncol Ther 2016;4:257–74.
34. Jänne PA, von Pawel J, Cohen RB, Crino L, Butts CA, Olson SS, et al. Multicenter, randomized, phase II trial of CI-1033, an irreversible pan-ERBB inhibitor, for previously treated advanced non small-cell lung cancer. J Clin Oncol 2007;25:3936–44.
35. Rixe O, Franco SX, Yardley DA, Johnston SR, Martin M, Arun BK, et al. A randomized, phase II, dose-finding study of the pan-ErbB receptor tyrosine-kinase inhibitor CI-1033 in patients with pretreated metastatic breast cancer. Cancer Chemother Pharmacol 2009;64:1139–48.
36. Campos S, Hamid O, Seiden MV, Oza A, Plante M, Potkul RK, et al. Multicenter, randomized phase II trial of oral CI-1033 for previously treated advanced ovarian cancer. J Clin Oncol 2005;23:5597–604.
37. Cassell A, Grandis JR. Investigational EGFR-targeted therapy in head and neck squamous cell carcinoma. Expert Opin Investig Drugs 2010;19:709–22.
38. Loeffler-Ragg J, Witsch-Baumgartner M, Tzankov A, Hilbe W, Schwentner I, Sprinzl GM, et al. Low incidence of mutations in EGFR kinase domain in Caucasian patients with head and neck squamous cell carcinoma. Eur J Cancer 2006;42:109–11.