Nawab *et al. BMC Cancer* (2025) 25:1333
https://doi.org/10.1186/s12885-025-14609-8

BMC Cancer

RESEARCH Open Access

Exploring the mutational spectrum of key kinase genes *PIK3CA*, *BRAF*, *EGFR*, *ALK* and *ROS1* in oral squamous cell carcinoma

Fouzia Nawab[1], Wafa Naeem[1], Sadia Fatima[1], Asif Ali[2,3,4*], Ali Talha Khalil[5*], Aamir Mehmood[6], Muhammad Fazeel[7], Hilal Ahmad[1], Mohammed Alorini[3], Muslim Khan[8], Ishtiaq Ahmad Khan[9], Muhammad Irfan[9] and Syed Ali Khurram[10*]

Abstract

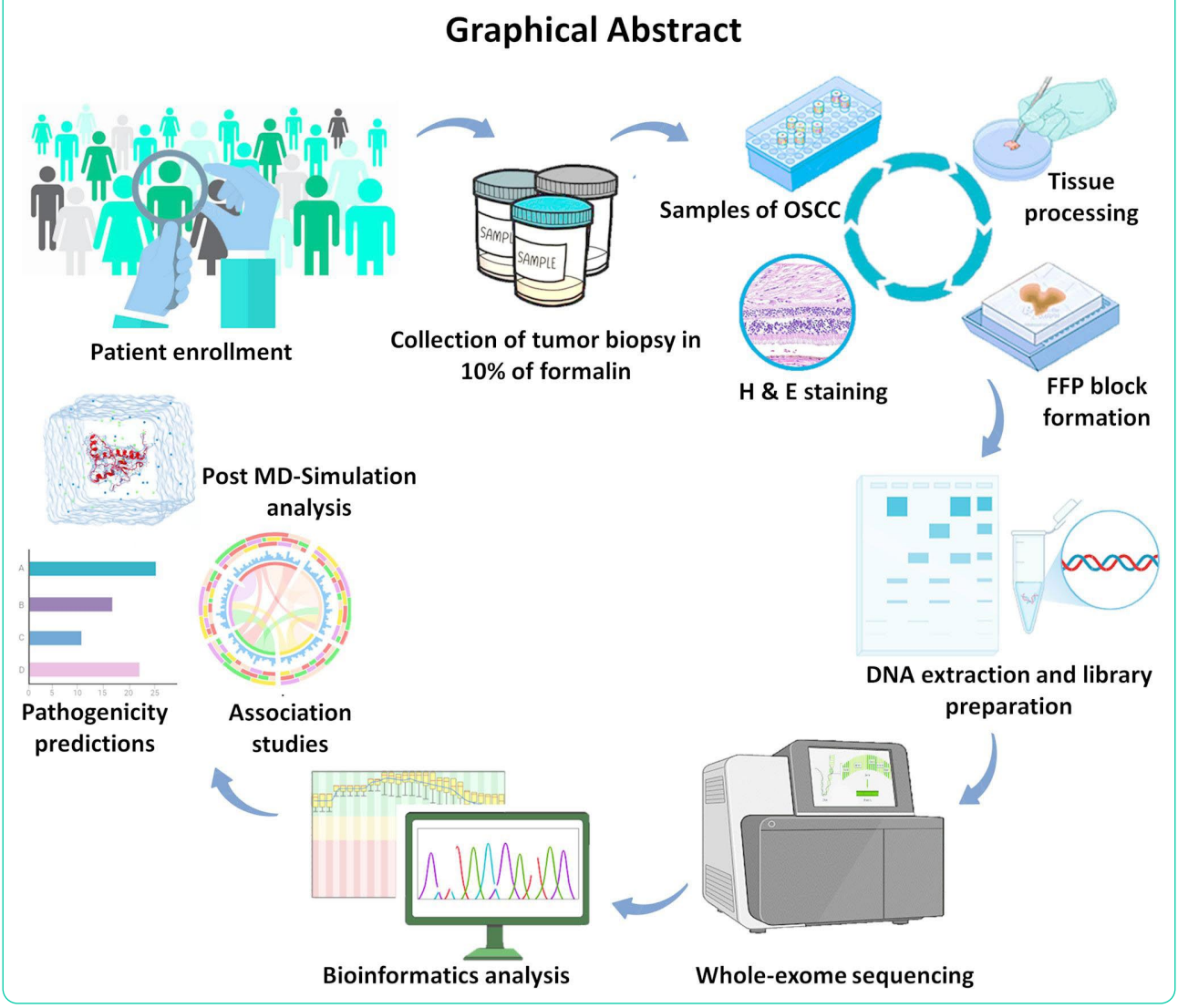
Oral Squamous Cell Carcinoma (OSCC) is the sixth most aggressive type of oral cancer. Mutations in cancer-driving genes such as protein kinase are well known in cancer progression. We selected candidate genes (*PIK3CA*, *BRAF*, *EGFR*, *ALK*, and *ROS1*) for mutations exploration in the OSCC patients belonging to Khyber Pakhtunkhwa (KP) through Next Generation-Whole Exome Sequencing (NG-WES) using Formalin Fixed Paraffin Embedded (FFPE) tissue blocks (27 tumor and 7 paired normal) for the 1st time followed by in-silico characterization. A total of 33 mutations were identified which constituted 28/33 (84.84%) SNVs, 4/33 (12.12%) frameshift deletions and 1/33 (3.03%) stop-gain mutation. While, of the 33 mutations, 12.6% (4/33) were novel and had not been previously reported in public mutation databases such as COSMIC or dbSNP. Among the total somatic mutations (24/33; 72.72%), 08/33 mutations were observed in multiple patients. Mutations of the *ALK* i.e. $ALK^{p.I1461V}$, $ALK^{p.K1491R}$ and $ALK^{p.D1529E}$ were found in 27/27, 21/27 and 20/27 tumor samples and hence can have potential biomarker applications. ISPRED-SEQ identified 07/33 interaction site mutations i.e. $EGFR^{p.R521K}$, $EGFR^{p.R831C}$, $ROS1^{p.S2229C}$, $ROS1^{p.E1902K}$, $ROS1^{p.K2228Q}$, $ROS1^{p.D2213N}$, and $ROS1^{p.P221S}$. SAAFEQ-SEQ predictions revealed that (28/29; 96.5%) SNVs destabilize protein except for $ROS1^{p.D2213N}$. ConSurf predictions indicated 17.3% (5/33) mutations (e.g., $ROS1^{p.N2240K}$ (score 9) and $ROS1^{p.L567V}$ (score 8), as highly conserved, likely disrupting kinase function and stability, unlike variable mutations with milder effects. MD simulations for interacting sites (IS) SNVs revealed structural deviations, with mutant proteins revealing larger gyration radius and fluctuations in root mean square deviation (RMSD) studies indicating a disrupted folding behavior. To conclude, we identified potential mutations on *ROS1* that can have potential biomarker applications, however, we recommend studies in large Pakhtun cohorts in KP, Pakistan.

Keywords Oral cancer, Kinase genes, NGS, Biomarker, Bioinformatics

*Correspondence:
Asif Ali
draliasif7@gmail.com
Ali Talha Khalil
alitalha.khalil@lrh.edu.pk
Syed Ali Khurram
s.a.khurram@sheffield.ac.uk

Full list of author information is available at the end of the article

Graphical Abstract



Introduction

Oral Squamous Cell Carcinoma (OSCC) is one of the most prevalent and aggressive types of cancer, constituting a significant global health burden [1]. It is the sixth most common oral cancer mainly arises from the tongue, buccal, palate, and floor of the mouth and offers a significant clinical challenge due to its high mortality and morbidity [2]. OSCC accounts for over 90% of all oral malignancies, with nearly 400,000 new cases reported annually [3]. Globally, OSCC has an incidence of 0.6 million and a mortality rate of 0.35 million, with South and Southeast Asia having exceptionally high incidence rates, according to the Global Cancer Observatory (2020) [4]. In Pakistan, OSCC is the most prevalent malignancy (incidence and mortalities) in terms of males, whereas it is second after breast cancer in females. Out of the total 185,748 new cases in 2022, 8.6% were of lip and oral cavity. Specifically in males, the prevalence is 12.3% and in females the prevalence is 5.3%. Figure 1, shows the most prevalent cancers in males and females of Pakistan and is derived from the Globcon report 2022 [5]. OSCC has a multifactorial etiology, with risk factors such as smoking, alcohol use, snuff dipping, and the use of smokeless tobacco and betel quid being notably common in South Asia [6]. According to estimates, in 2022, ~ 120,000 new cases of oral cancers were attributed to the use of smokeless tobacco and areca nut with a high proportion especially in South-Central Asia and LMIC's [7]. The five year survival rate in oral cancers is < 50%, with comparatively better outcomes in women [8].

Surgery is the most practiced option but leads to significant deformity, usually causing the inability of mouth

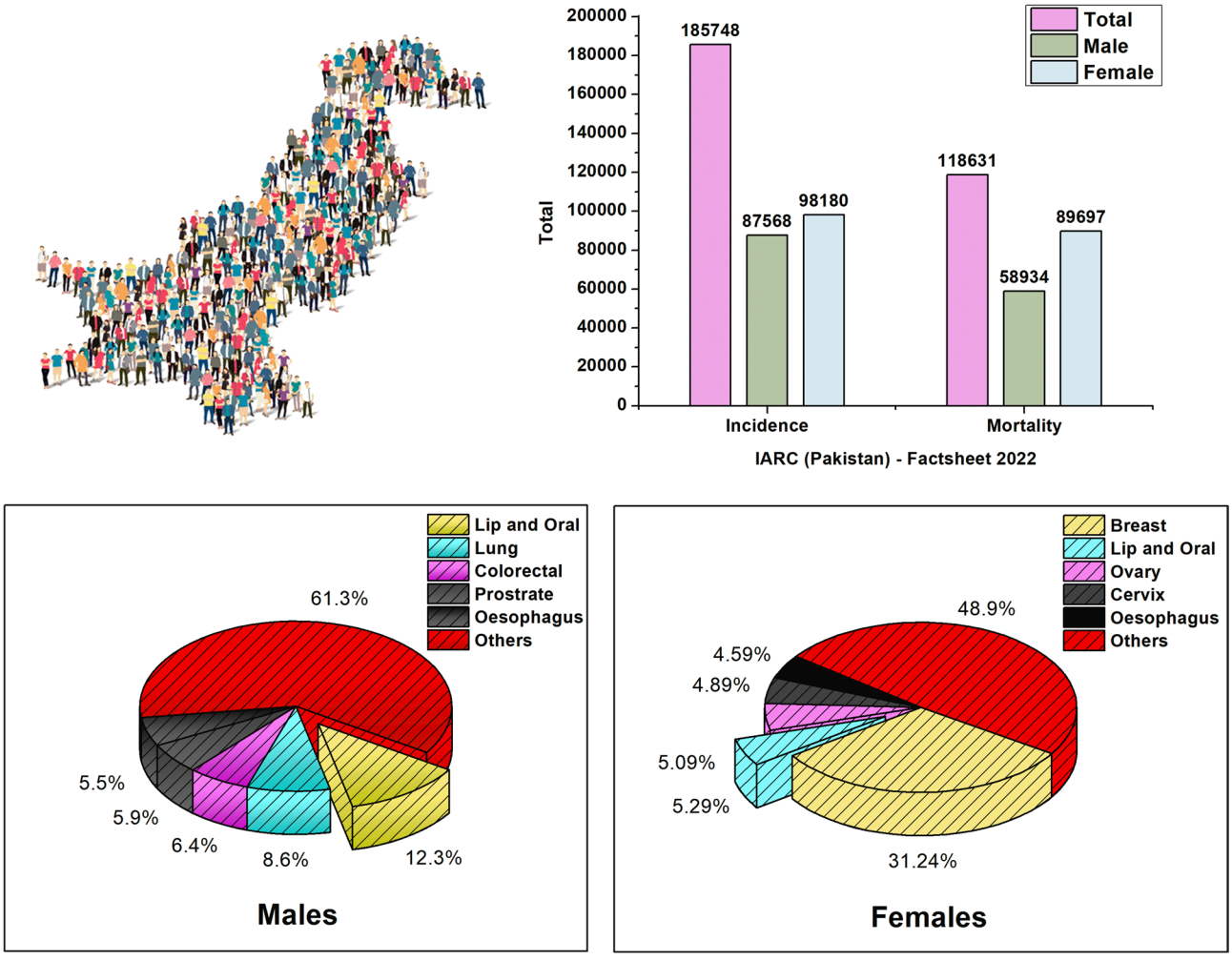


Fig. 1 Incidence, mortality and number of new cases in Pakistan according to the country factsheet from Global Cancer Observatory (Globcon)

functions, psychosocial stress and extensive rehabilitation. The application of chemotherapy and radiotherapy is limited because of intolerance and toxicity. Immunological treatments are still in their infancy. Given the limitations of current treatments and the high morbidity of OSCC, identifying driver mutations as biomarkers is critical for advancing early diagnosis and targeted therapeutic strategies [8, 9].

Despite advancements in treatment strategies, the prognosis for OSCC remains poor. High rates of recurrence and metastasis, contributes to its elevated mortality and morbidity. Therefore, understanding the molecular processes that drive OSCC growth is crucial [10]. Recently, the conception of the molecular mechanisms underlying OSCC has expanded significantly, with a growing emphasis on the role of specific genetic mutations in its pathogenesis [11]. Kinase genes, such as *PIK3CA*, *BRAF*, *EGFR*, *ALK*, and *ROS1*, are critical signaling mediators regulating cellular processes like proliferation and apoptosis, and their dysregulation drives OSCC oncogenesis [12, 13].

Despite the clinical relevance of kinase genes mutations in cancer, research on their profiles in the Pakistani population, particularly in Khyber Pakhtunkhwa, remains largely unexplored. By using whole-exome sequencing data, this study focuses on profiling genetic mutations and their correlation with histopathological and sociodemographic variables in five key kinase genes: *PIK3CA*, *BRAF*, *EGFR*, *ROS1*, and *ALK* in patients with OSCC. Furthermore, through in-silico approaches including, computational protein modeling and bioinformatics analyses, we aim to identify the functional consequences of these mutations and assess their prognostic and therapeutic potential as biomarkers. Identifying these genetic drivers and mapping their mutational profiling in OSCC will be pivotal for tailoring precision medicine approaches, thus contributing to more individualized and effective cancer management.

Materials and methods

Inclusion/exclusion criteria

Patients of both genders and all age groups who were clinically and histologically confirmed to have OSCC (Stage I-IV) were included. While patients who experienced tumor recurrence, received alternative treatments such as chemotherapy or radiotherapy, or had other types of cancers were excluded from the study.

Subject characteristics and sample processing

A total of 27 tumor tissue samples and 7 paired normal samples were selected for this study. The tumor tissues were obtained from patients diagnosed with Oral Squamous Cell Carcinoma (OSCC) at Hayatabad Medical Complex (HMC) and Khyber College of Dentistry (KCD), Peshawar, KP, Pakistan. A pathologist identified the paired normal tissues. The samples were obtained in 10% buffered formalin and transported to the Khyber Medical University (KMU) laboratory, where their formalin-fixed paraffin-embedded (FFPE) blocks were made for further processing. Moreover, Hematoxylin and eosin (H&E) staining was performed on the sections from these blocks for histopathological analysis. All the steps and procedures were conducted in line with the ethical guidelines of the Declaration of Helsinki. Approval for the study was obtained from the research ethics committee of Khyber Medical University, Peshawar (Reference number *Dir/Ethics/KMU/2020/17;* dated *29/01/2020*). Written informed consent was obtained from all patients prior to sample collection. Based on inclusion criteria, patients of all ages and both genders who were clinically and histologically confirmed OSCC patients were included.

DNA extraction and quality assessment

Genomic DNA was isolated from the formalin fixed paraffin embedded tumors using QIAmp DNA FFPE Tissue Kit (Catalog No. 56404) by following the kit protocols. The quality of isolated DNA was evaluated by 2% agarose gel electrophoresis while the concentration of DNA was assessed using high sensitivity (HS) double stranded DNA Qubit kit by Qubit Fluorometer 2.0 (Thermo Fisher). The good quality isolated DNA was stored at –20 ºC till further process.

DNA library preparation and sequencing

High quality isolated DNA (200 ng) was processed for library preparation using Illumina DNA Prep with Exome 2.5 Enrichment Kit (Illumina), following the manufacturer's protocol. Briefly, genomic DNA was fragmented into about 300 bases using transposes followed by adapter ligation at the blunt end followed by few PCR cycle amplification. The amplified products were purified by Agencourt Ampure magnetic beads and the concentration of the purified library was evaluated by HS dsDNA Qubit kit. DNA fragments containing exonic regions were captured by hybridizing with specific coding oligos (CEX). The captured coding fragments were purified followed by enrichment with few PCR cycle amplification. The enriched library was purified using Agencourt Ampure magnetic beads. Size distribution of the library was evaluated by 2% Agarose gel electrophoresis and the concentration was assessed by Qubit Fluorometer. The library was normalized and diluted to 20 pmole using hybridization buffer (HT1). The library was further diluted to 1.8 pmole for high through put sequencing using high through put NextSeq500/550 sequencing cartridge. The paired end sequencing was performed using 2 × 150 cycle flowcell chemistry on NextSeq500 Illumina platform.

Analysis and annotation of sequence data

To analyze sequence data, the raw sequencing data was demultiplexed and converted to fastq using bcl2fastq. The quality of sequencing reads was evaluated by fastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Adapter sequences and the poor-quality reads were removed using fastp tool (http://opengene.org/fastp/fastp). Good quality sequencing data was aligned with hg38 reference genome (hg38-UCSC) using Burrows – Wheeler aligner (BWA-mem) tool. The sequenced aligned reads in the bam format were sorted and the PCR duplicated reads were marked and removed using PICARD tool. Base recalibration and haplotype calling was carried out by standard GATK pipeline. Variants with a quality score (QUAL) greater than 30, read depth (DP) less than 20, genotyping quality (GQ) above 20, and minor allele frequency (MAF) below 0.01 in the gnomAD v4.1 were used as thresholds for variant calling. The data was processed using the ANNOVAR tool to generate a comprehensive CSV file containing all relevant information. Further filtration of the data was carried out with the help of R program.

Bioinformatics analysis

To predict the pathogenicity of all the nonsynonymous SNVs, five different tools SIFT, Polyphen-2, Mutation Taster, Mutation Assessor, PROVEAN, FATHMM were used. ISPRED-SEQ (https://ispredws.biocomp.unibo.it/sequence/) was used to predict the mutations of the inter action sites. SAAFEQ-SEQ (http://compbio.clemson.edu/lab/) was used to predict the effect of mutations on pr otein stability. For predicting the evolutionary conservation of the mutated residue position, and for determining the other attributes i.e. functional, structural, buried and exposed nature, "ConSurf" tool (https://consurf.tau.ac.il/) was utilized.

Molecular modeling and mutation mapping

"maftool" in R-Studio was used to create the lollipop plots in order to show location and distribution of different mutations across *PIK3CA*, *BRAF*, *EGFR*, *ALK* and *ROS1* genes in the recruited oral squamous cell carcinoma (OSCC) cancer patients [14]. For molecular modeling, the proteins with mutations on interaction site residues were modeled using "Swiss-Model" (https://swissmodel.expasy.org/) and then were superimposed and visualized using "PyMOL" software [15]. Furthermore, interaction networks for its proteins were analyzed using STRING and GeneMANIA databases. The mutational spectrum of these kinase genes were associated with clinicopathological and sociodemographic determinants in OSCC patients.

Molecular dynamics (MD) simulation

The selected interaction site mutations were subjected to molecular dynamics (MD) simulations using GROMACS 5.1. Firstly, the 3D structures of the proteins were prepared using Swiss-Model, and the resultant Pdb files were used for simulations. Next, a topology file was generated, incorporating the appropriate force field (OPLS-AA/L all-atom) parameters and molecular types to define the molecular structure and interactions within the system. Solvation was performed, for which the solvated system was prepared by embedding the protein in a box with the solvent model SPC216.gro, providing a realistic aqueous environment. Further, to neutralize the system, counter ions were added to offset any residual net charge in the system. We then subjected the system to an energy minimization step, aiming to relieve unfavorable contacts and ensure a stable starting configuration. This initial minimization was essential to eliminate steric clashes and allow the system to begin from a low-energy state [16]. Equilibration was performed in two sequential phases to stabilize both the temperature and pressure of the system. First, an NVT (constant number of particles, volume, and temperature) equilibration phase was applied to stabilize the temperature, ensuring a consistent thermal environment. This was followed by an NPT (constant number of particles, pressure, and temperature) equilibration phase to bring the system to equilibrium at a constant pressure and density. These equilibration steps were crucial to prepare the system for the subsequent production MD simulation phase, ensuring that the system parameters were stabilized under physiological conditions. The final step was to run apo simulations lasting 50ns.

Furthermore, during the production of MD simulation, a range of structural and dynamic analyses were conducted over time to evaluate differences between mutant (MT) and wild-type (WT) structures. Root-mean-square deviation (RMSD) was calculated to monitor the overall stability of the protein backbone throughout the simulation. The root-mean-square fluctuation (RMSF) was determined to assess local structural variations, allowing us to identify residues with increased flexibility due to mutation. To evaluate the protein's folding and compactness, the radius of gyration (Rg) was tracked over time, providing a measure of the protein's overall structural stability [17]. Additionally, GROMACS energy parameters (including; temperature, pressure, density, and potential) were recorded and all the results were visualized using ORIGIN PRO 2024b. The wild type and mutant protein models were evaluated for further validation by making Ramachandran plots using the PROCHECK server.

Results

Sample characteristics

A total of 27 patients that met the inclusion criteria participated in the study. Our findings revealed a higher prevalence of OSCC cases in male (19/27; 70.4%) as compared to female (8/27; 29.6%) population. Regarding age, 16 patients were above 56 years (16/27; 59.3%), while 11 patients were under 56 years (11/27; 40.7%). Anatomically, the tumors were distributed across several sites, including the tongue (10/27; 37% cases), lip (5/27; 18.5% cases), buccal mucosa (4/27; 14.8% cases), and other areas i.e. Mandible, Oral cavity, palate and floor of mouth (8/27; 29.6% cases). Histopathological analysis revealed that 14 cases (51.8%) were well differentiated, while 13 (48.1%) were moderately differentiated. Furthermore, with respect to sociodemographic factors, the analysis of tobacco consumption revealed that 10 patients (10/27; 37%) were non- tobacco users, 15 (15/27; 55.5%) used naswar, and 2 (2/27; 7.4%) were smokers. Additionally, 12/27; 44.44% patients had a favorable family history of cancer, while 15/27; 55.55% patients did not have a positive family history. In terms of dental issues, 8 (8/27; 29.6%) patients reported a history of dental problems, whereas 19/27 (70.37%) did not have any history of dental issues.

Mutational profiling

For mutation analysis of selected genes; *PIK3CA*, *BRAF*, *EGFR*, *ALK*, and *ROS1*, the whole-exome sequencing (WES) data were analyzed using several mutation databases, including the COSMIC database and dbSNP. Variants found in both tumor and paired normal tissue samples were classified as germline mutations, whereas the gene variants found only in tumor tissue samples were labelled as somatic mutations [18]. The overall spectrum of mutations identified in the selected genes is summarized in Table 1. These included 28 nonsynonymous SNVs (28/33; 84.84%), 4 frameshift deletions (4/33; 12.12%), and 1 stop-gain mutation (1/33; 3.03%) (Fig. 2C). Among the total 33 mutations, 24 (72.72%)

Table 1 Summary of the mutational landscape of *PIK3CA*, *BRAF*, *EGFR*, *ALK* and *ROS1*

Patients	Mutation Type	Reported	Status	Exon	Mutation		Pathogenicity					
					Nucleotide Change	Effect on Protein	SIFT Pred	Poly-phen- 2 Pred	Muta-tion Taster Pred	Mutation Assessor Pred	PROVEAN Pred	FATHMM Pred
					PIK3CA							
2	nonsynonymous SNV	dbSNP	Somatic	7	c.1173 A > G	p.I391M	T	B	p	L	N	T
1	nonsynonymous SNV	-	Somatic	10	c.1585G > C	p.E529Q	T	B	D	L	N	T
1	nonsynonymous SNV	-	Somatic	10	c.1604 C > T	p.S535F	T	B	D	L	N	T
1	nonsynonymous SNV	COSMIC/dbSNP	Somatic	10	c.1624G > A	p.E542K	D	D	D	L	N	T
1	nonsynonymous SNV	COSMIC/dbSNP	Somatic	10	c.1634 A > G	p.E545G	D	D	D	M	D	T
1	frameshift deletion	Novel	Somatic	10	c.1549_1552del	p.A518Efs*5	-	-	-	-	-	-
					BRAF							
1	nonsynonymous SNV	dbSNP	Germline	7	c.757 C > G	p.P253A	D	B	D	M	N	T
1	Frameshift deletion	Novel	Somatic	4	c.561delG	p.M187Ifs*4	-	-	-	-	-	-
1	Frameshift deletion	Novel	Somatic	12	c.1575delG	p.T526Lfs*31	-	-	-	-	-	-
					EGFR							
12	nonsynonymous SNV	COSMIC/dbSNP	Germline	13	c.1562G > A	p.R521K	T	B	P	N	N	T
2	nonsynonymous SNV	COSMIC/dbSNP	Somatic	21	c.2491 C > T	p.R831C	D	D	D	L	D	D
					ALK							
2	nonsynonymous SNV	dbSNP	Somatic	7	c.1427T > C	p.V476A	T	B	D	N	N	T
1	nonsynonymous SNV	-	Somatic	11	c.1917 C > A	p.S639R	T	B	D	L	N	T
1	frameshift deletion	Novel	Somatic	11	c.1919delG	p.G640Efs*25	-	-	-	-	-	-
1	nonsynonymous SNV	dbSNP	Germline	18	c.3035 C > T	p.T1012M	T	B	D	N	N	T
20	nonsynonymous SNV	dbSNP	Germline	29	c.4587 C > G	p.D1529E	T	B	P	N	N	T
21	nonsynonymous SNV	dbSNP	Germline	29	c.4472 A > G	p.K1491R	T	B	P	L	N	T
27	nonsynonymous SNV	COSMIC/dbSNP	Germline	29	c.4381 A > G	p.I1461V	T	B	P	N	N	T
					ROS1							
1	nonsynonymous SNV	dbSNP	Somatic	12	c.1611 A > G	p.I537M	T	B	P	M	N	T
1	nonsynonymous SNV	dbSNP	Somatic	12	c.1699 C > G	p.L567V	D	D	D	L	N	D
1	nonsynonymous SNV	-	Somatic	17	c.2530 C > T	p.L844F	T	B	N	N	N	D
1	nonsynonymous SNV	-	Somatic	30	c.5068G > A	p.E1690K	D	D	D	M	N	T
8	nonsynonymous SNV	dbSNP	Germline	5	c.433 A > C	p.T145P	T	D	P	M	N	T
1	StopGain	dbSNP	Somatic	17	c.2548 C > T	p.Q850X	-	-	A	-	-	-
2	nonsynonymous SNV	dbSNP	Somatic	21	c.3326 C > T	p.S1109L	D	B	N	M	D	T
1	nonsynonymous SNV	-	Germline	34	c.5617G > A	p.V1873I	T	B	N	L	N	T
8	nonsynonymous SNV	dbSNP	Somatic	42	c.6686 C > G	p.S2229C	T	B	P	N	N	T
3	nonsynonymous SNV	dbSNP	Somatic	6	c.500G > A	p.R167Q	D	B	P	L	D	T
1	nonsynonymous SNV	dbSNP	Somatic	35	c.5704G > A	p.E1902K	D	B	D	M	N	T
8	nonsynonymous SNV	dbSNP	Somatic	42	c.6682 A > C	p.K2228Q	T	B	P	N	N	T

Table 1 (continued)

Patients	Mutation Type	Reported	Status	Exon	Mutation		Pathogenicity					
					Nucleotide Change	Effect on Protein	SIFT Pred	Poly-phen-2 Pred	Muta-tion Taster Pred	Mutation Assessor Pred	PROVEAN Pred	FATHMM Pred
8	nonsynonymous SNV	dbSNP	Somatic	42	c.6637G>A	p.D2213N	T	B	P	L	N	D
2	nonsynonymous SNV	dbSNP	Germline	42	c.6720T>G	p.N2240K	D	D	D	M	N	T
1	nonsynonymous SNV	-	Somatic	7	c.661 C>T	p.P221S	D	D	D	L	D	T

Legends: "SIFT: Sorting Intolerant From Tolerant"; D: Deleterious; T: Tolerated;
"PolyPhen-2: Polymorphism Phenotyping version" 2 D: Probably damaging; P: Possibly damaging; B Benign;
"Mutation Taster": A: Disease causing automatic; D: Disease causing; N: Polymorphism; P: Polymorphism automatic;
"Mutation Assessor": L: Low; M: Medium; N: Neutral;
"PROVEAN": Protein Variation Effect Analyzer"; N: Neutral; D: Deleterious;
"FATHMM": Functional Analysis Through Hidden Markov Models; D: Deleterious; T: Tolerated

were somatic and 9 (27.27%) were germline mutations, as determined by comparison with paired normal tissues (27 tumor tissue samples and 7 paired normal tissues) (Fig. 2B). Notably, 4 out of the 33 mutations (12.1%) are novel and reported here for the first time. The gene-wise incidence of somatic mutations on *PIK3CA*, *BRAF*, *EGFR*, *ALK*, and *ROS1* were 25% (6/24), 8.33% (2/24), 4.16% (1/24), 12.5% (3/24) and 50% (12/24). No germline mutation was identified on *PIK3CA*. On *BRAF*, *EGFR*, *ALK*, and *ROS1* the germline mutations were 11.1% (1/9), 11.1% (1/9), 44.4% (4/9), and 44.4% (3/9) respectively. All the six mutations on *PIK3CA* were somatic (5 SNVs and 1 Frameshift Deletion). One novel frameshift deletion i.e. $PIK3CA^{p.A518Efs*5}$ was identified on *PIK3CA*. The *BRAF* gene had three mutations, including one nonsynonymous SNV; $BRAF^{p.P253A}$ (germline) and two novel frameshift deletion mutations i.e. $BRAF^{p.M187Ifs*4}$ and $BRAF^{p.T526Lfs*31}$ (somatic). *EGFR* gene had two nonsynonymous mutations (including 1 somatic and 1 germline). The *ALK* gene had seven mutations, consisting of six nonsynonymous SNVs and one frameshift deletion as a novel mutation; $ALK^{p.G640Efs*25}$. Moreover, *ROS1* gene had a total of 15 mutations including 14 nonsynonymous and one stop gain somatic mutation. Mutations on *ALK* i.e. $ALK^{p.I1461V}$ was found in all patients (27/27; 100%), whereas, $ALK^{p.K1491R}$ (21/27; 77.77%) and $ALK^{p.D1529E}$ (20/27; 74.07%) were also found in majority of the population indicating their potential biomarker applications in local population. Similarly, mutations on *ROS1*, such as $ROS1^{p.K2228Q,}ROS1^{p.D2213N,}ROS1^{p.S2229C,}ROS1^{p.T145P}$ was found to be recurring in 8/27 (29.62%) patients and can be further evaluated in larger cohorts for biomarkers applications (Fig. 2D).

The lollipop plots illustrate the location and sequence of the most notable genetic alterations are indicated in Fig. 3A–E.

The inset of Fig. 4A – E indicates the frequency of mutations on exons of *PIK3CA*, *BRAF*, *EGFR*, *ALK* and *ROS1* to determine the exon locations where mutations occurred most frequently. For *PIK3CA*, *ALK* and *ROS1*, the highest mutation frequencies were observed on exon 10, exon 29, and exon 42, respectively. However, for *BRAF* and *EGFR* gene, the mutations were found on exon 4, 7, 12, 13 and 21.

Prediction of tolerated and deleterious SNVs

The pathogenic impact of mutations identified in the *PIK3CA*, *BRAF*, *EGFR*, *ALK and ROS1* genes was predicted using a combination of in-silico tools, including SIFT, PolyPhen-2, Mutation Taster, Mutation Assessor, PROVEAN and FATHMM. These tools assess whether the identified mutations are likely to be deleterious or benign based on evolutionary conservation, protein structure, and biochemical properties [19]. The predicted

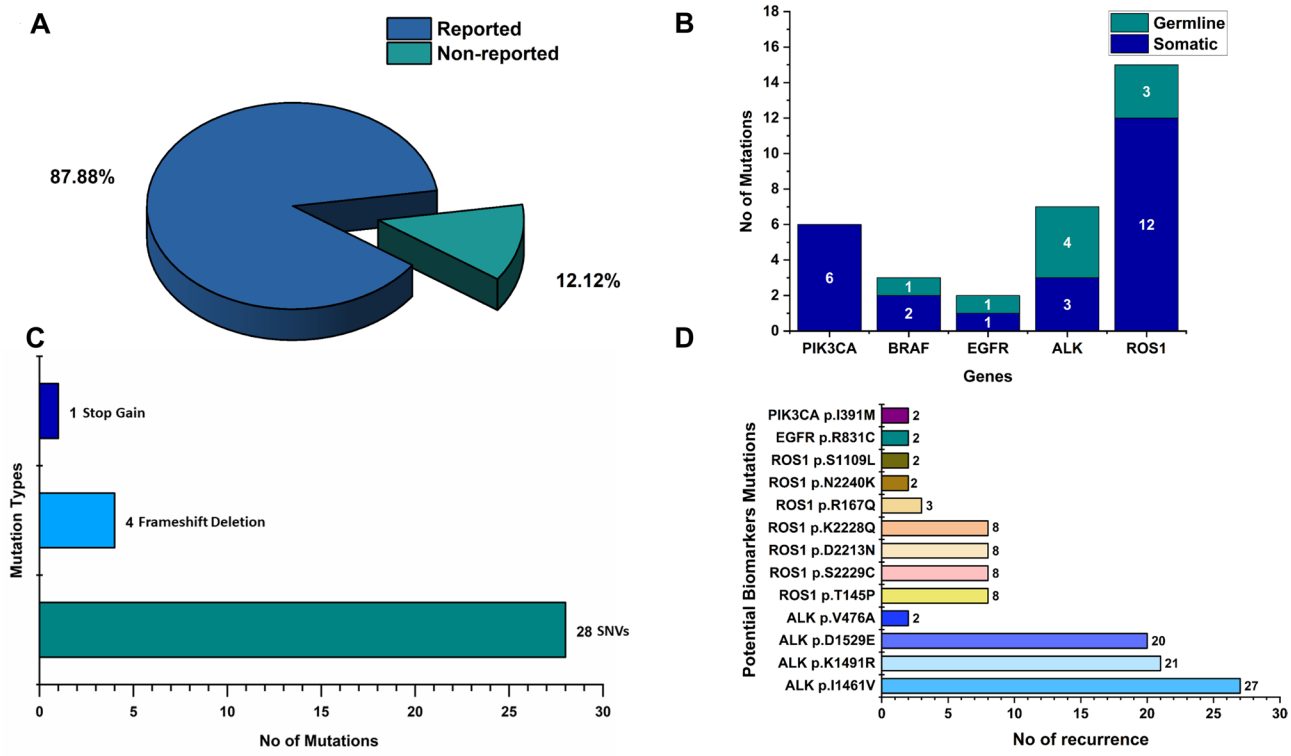


Fig. 2 Mutational profile of selected key kinase genes (*PIK3CA*, *ROS1*, *ALK*, *EGFR*, *BRAF*) **A** Percentage of novel mutations and reported mutations from the data obtained using the COSMIC/dbSNP databases and literature survey; **B** Frequency of germline and somatic mutations; **C** Mutation type in the selected genes on basis of mutation rate; **D** Recurring mutations having Biomarker Potential

pathogenicity results of mutations in the selected genes are summarized in Figure S1 (A-F). According to SIFT prediction, overall, 11/33 (33.33%) predictions were considered as deleterious and 17/33 (51.51%) are tolerated. However, 5/33 (15.15%) mutations do not revealed any results in SIFT predictions. On the other hand, the SIFT prediction for *PIK3CA*, *BRAF*, *EGFR*, *ROS1* revealed a total of 2/6 (33.3%), 1/3 (33.3%), 1/2 (50%), and 7/15 (46.7%) mutations as deleterious respectively. While for *ALK* mutations, no such mutation was categorized as deleterious. The Polyphen-2 predictions for *PIK3CA*, *EGFR* and *ROS1* revealed 2/6 (33.3%), 1/2 (50%) and 5/15 (33.3%) as probably damaging and 3/6 (50%), 1/2 (50%) and 9/15 (60%) as benign mutations. However, 1/15; 6.66% *(ROS1*), 1/7; 14.2% (*ALK1*), 1/6; 16.6% (*PIK3CA*) and 2/3; 6.6% (*BRAF*) had no prediction score according to polyphen-2 (Figure S1-B). The prediction from Mutation Taster database revealed 4/6 (66.67%), 1/3 (33.3%), 1/2 (50%), 3/7 (42.8%) and 5/15 (33.3%) mutations as disease causing in all the five genes. However, 1/7; 14.2% (*ALK1*), 1/6; 16.6% (*PIK3CA*) and 2/3; 66.6% (*BRAF*) were not predicted according to Mutation Taster (Figure S1-C). Similarly, based on the medium and low impact categories observed according to Mutation assessor, for *ROS1*, *PIK3CA*, and *BRAF* 6/15 (40%), 1/6 (16.6%) and 1/3 (33.3%) of the mutations were classified as medium impact, 5/15; 33.3% (*ROS1*), 2/7; 28.6% (*ALK*) and 4/6; 66.6% (*PIK3CA*) as low impact mutations. While, 1/15; 6.66% (*ROS1*), 2/3; 66.6% (*BRAF*), 1/7; 14.3% (*ALK*) and 1/6; 16.6% (*PIK3CA*) as non-predicted mutations according to Mutation Assessor database (Figure S1-D). According to the PROVEAN analysis, out of 33, 5 mutations (5/33; 15.15%) were predicted to be 'deleterious' in all the selected genes and 23 (23/33; 69.69%) mutations were predicted to be 'neutral' as shown in Figure S1-F. Similarly, according to FATHMM results, overall, 24 out of 33 (72.72%) mutations were considered as tolerated, while 4 out of 33 (12.12%) mutations were considered as deleterious mutations. Table S1 shows the unpredicted and predicted mutations from the various databases.

Prediction of protein structural stability and interacting site mutations

To further scrutinize these mutations, another tool SAAFEQ-SEQ predictions was used that relate to the effect of the SNVs on protein stability. Our results revealed the destabilizing effect predicted for all single nucleotide variants (SNVs) in the dataset, except for the $ROS1^{D2213N}$ mutation, which showed a stabilizing effect (Table S2). Additionally, we employed ISPRED-SEQ (an interaction site prediction tool) in order to determine SNVs (single nucleotide variations) of the selected gene

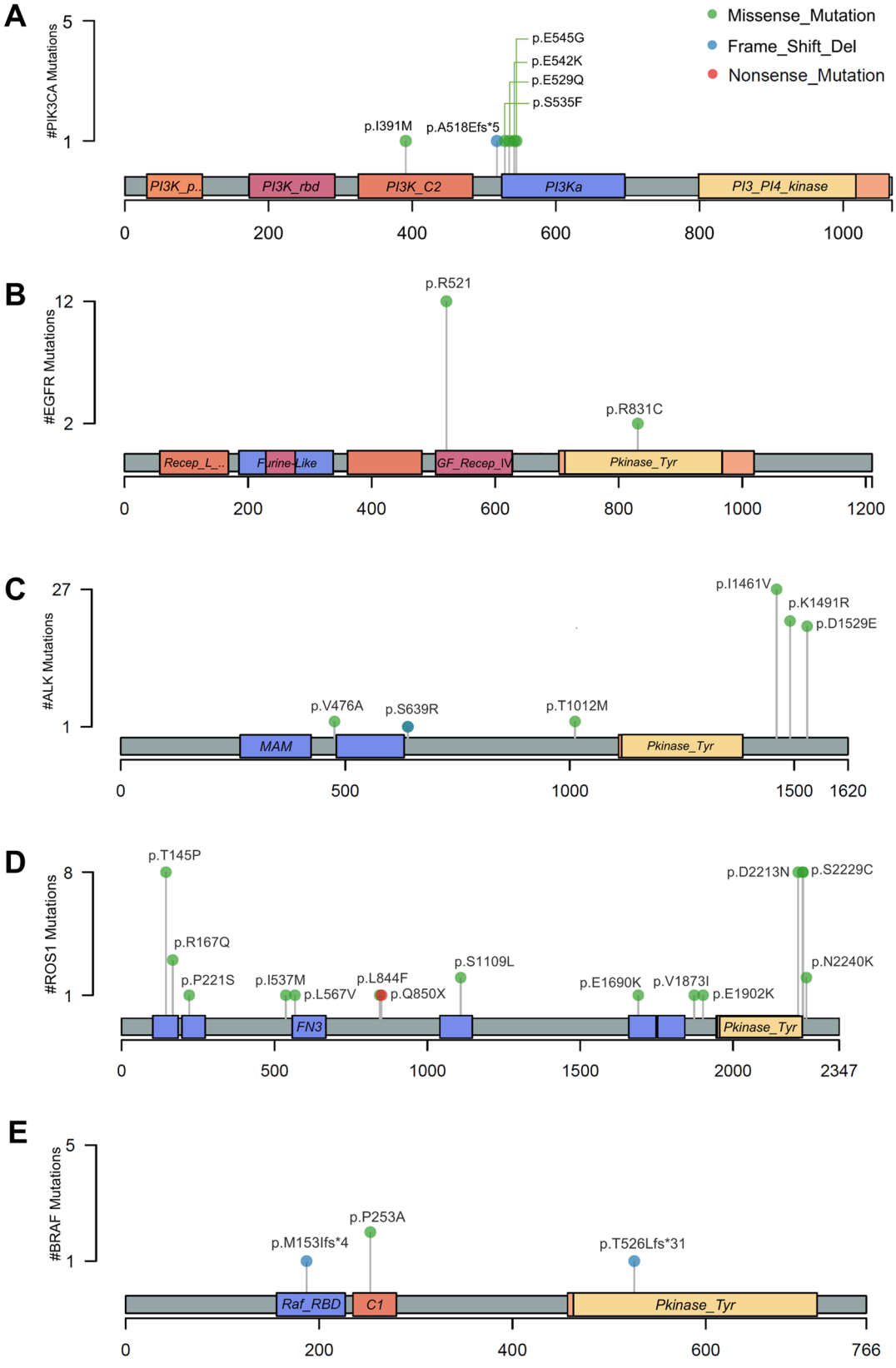


Fig. 3 Lollipop plot presenting the distribution of different mutations in *PIK3CA*, *EGFR*, *ALK*, *ROS1* and *BRAF* genes. The plots were created by the maftools in R-Studio

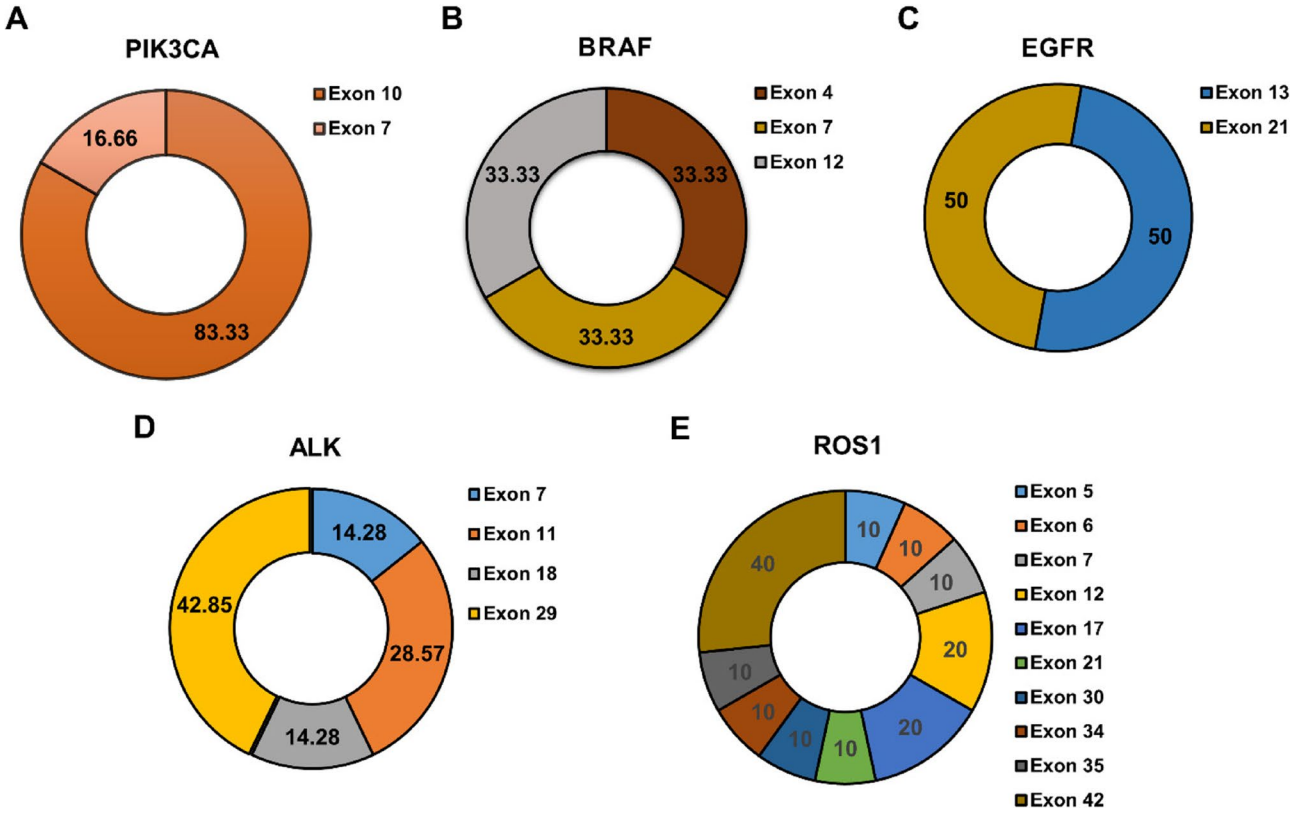


Fig. 4 Frequency (%) of Mutations on Exons of **A** *PIK3CA*, **B** *BRAF*, **C** *EGFR*, **D** *ALK*, **E** *ROS1*

at the interaction sites. From our results we found that, both mutations of *EGFR* ($EGFR^{R521K}$, $EGFR^{R831C}$) were interaction site (IS) mutations. Similarly, in *ROS1* 5/15 (33.3%) mutations ($ROS1^{S2229C}$, $ROS1^{E1902K}$, $ROS1^{K2228Q}$, $ROS1^{D2213N}$, $ROS1^{P221S}$) were present at the interacting sites. All mutations harbored by *BRAF, ALK and PIK3CA* were identified as non-IS mutations, with probability scores less than 0.5 (Table S3). The selected IS mutations in *EGFR* and *ROS1* were visualized and superimposed in PyMOL as shown in Fig. 5.

ConSurf predictions

The evolutionary conservation of amino acid residues in the selected genes were analyzed using the ConSurf server. Among the identified mutations in all genes, overall, 12/33; 36.36% were identified with highly and moderately conserved status having an exposed/buried nature, indicating their potential involvement in structural and functional alterations of the protein. Additionally, 12/33; 36.36% mutations were identified as highly and moderately variable status with exposed/buried nature residues, having a varied conservation scores, while 4/33; 12.12% of the mutation have an average conservation score of 5 with exposed or buried nature. Five mutations i.e. $ROS1^{p.N2240K}$, $ROS1^{p.L567V}$, $ROS1^{p.R167Q}$, $ROS1^{p.E1690K}$, $ROS1^{p.T145P}$ in *ROS1* gene were found to be located in highly conserved regions, each having the highest conservation scores of 9 and 8. $ROS1^{p.N2240K}ROS1^{p.E1690K}$ and $ROS1^{p.R167Q}$ were of exposed and functional nature while $ROS1^{p.L567V}$*and* $ROS1^{p.T145P}$ were predicted as residues with buried nature. For *EGFR*, the IS mutation $EGFR^{p.R521K}$ was predicted to be in an exposed region (moderately variable) status with a conservation score of 3. In contrast, the $EGFR^{p.R831C}$ mutation was identified to be of moderately conserved status with buried nature (conservation score 6). The results are summarized in Fig. 6 and Table S4.

MD simulations

In our study, we utilized molecular dynamics (MD) simulations to compare the structural stability and dynamics of seven different interacting sites mutations, analyzing key parameters such as root mean square deviation (RMSD), root mean square fluctuation (RMSF), Radius of gyration (Rg), temperature, pressure, density and potential. Over the duration of 50 ns, by analyzing the structural compactness (Rg) plots for the selected mutations (Figs. 7, 8, 9, 10, 11, 12 and 13A), compared to the wild type of proteins, all the mutant types have showed a greater Rg values, which generally illustrates abnormal folding of protein and decreased stability. Among the examined mutations, for $EGFR^{p.R521K}$, contrary to its

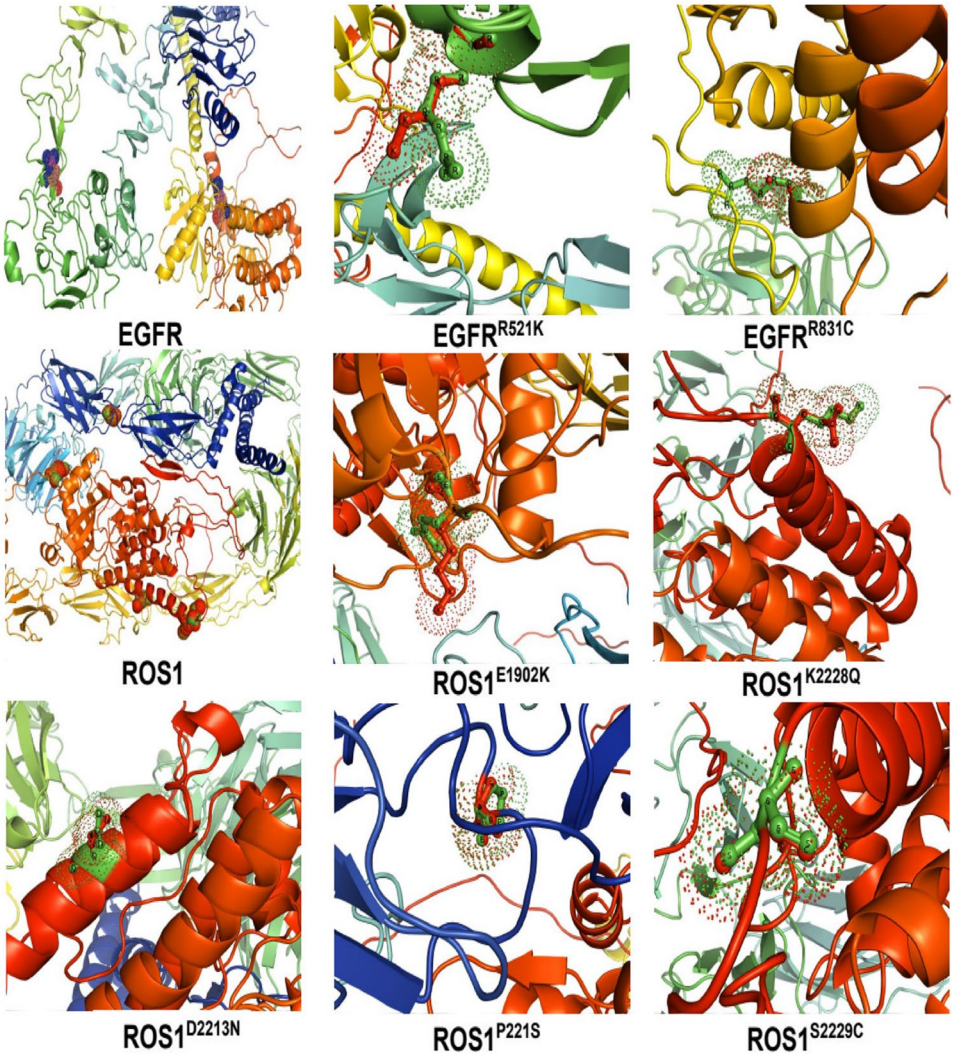


Fig. 5 *EGFR* and *ROS1* interacting site mutations visualization and superimposition in PyMOL

wild type (WT), the mutant type (MT) shows a high Rg value of 2.7 nm in the first 20 ns with significant variations before dropping abruptly to 2.2 nm and remains fluctuating between 2.2 and 1.8 nm until the simulation's end. Major differences were seen in *ROS1* IS variants ($ROS1^{p.S2229C}$, $ROS1^{p.P221S}$, $ROS1^{p.D2213N}$), in which throughout the simulations (50 ns), the Rg values stays higher as compared to WT proteins. The difference between the average radius of gyration values is shown in Fig. 14 for all the IS mutations.

Major structural deviations were observed while comparing the mutant and wild type proteins during RMSD results. For $EGFR^{p.R521K}$, a major deviation was observed at 25 ns and 33 ns compared to the wild type as shown in Fig. 7B. Major fluctuations were seen in $EGFR^{p.R831C}$ from 10 ns to 20 ns and a significant change in the trajectory after 20 ns as indicated in Fig. 8B. Similarly, for *ROS1*, the mutations in targeted structures showed varied deviations relative to their WT proteins (Figs. 9B, 10, 11, 12 and 13B). $ROS1^{p.S2229C}$ exhibited major deviations after 7 ns. (Fig. 9B). $ROS1^{p.E1902K}$ exhibited major deviations at 3 ns and minor fluctuations after 12ns (Fig. 10B). For $ROS1^{p.K2228Q}$, the fluctuations of mutant protein starts at 2 ns then drops at 5ns and again deviates upward at 16ns (Fig. 11B). Similarly, $ROS1^{p.P221S}$ mutant protein also showed major deviations as compared to WT, starting from 0 ns until the simulations end (Fig. 12B). Moreover, $ROS1^{p.D2213N}$ showed greater RMS deviations at different points as compared WT (Fig. 13B).

To study alterations in residue dynamics caused by selected IS mutations, RMSF values were computed. For *ROS1*, until the simulation's last frame, all the MTs had higher RMSF scores than the WTs, with considerable variations, especially around the changed residues (Figs. 9C, 10, 11, 12 and 13C). Similarly, in case of *EGFR*, Fig. 7 C also revealed notable changes in amino acid variations compared to the WT. For $EGFR^{p.R831C}$, overall, both the WT and mutant have a competing RMSF

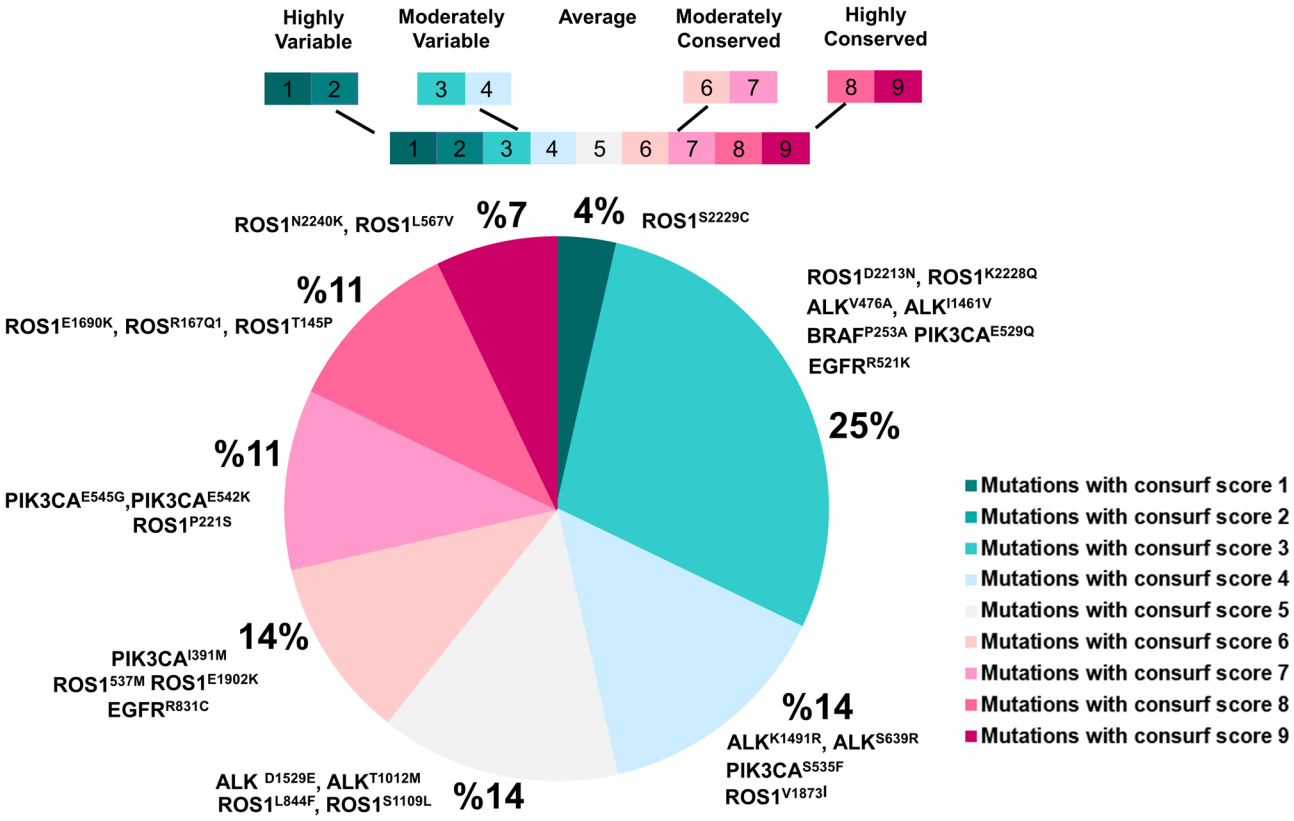


Fig. 6 ConSurf Scores Predictions of *PIK3CA*, *BRAF*, *EGFR*, *ALK* and *ROS1* Mutations

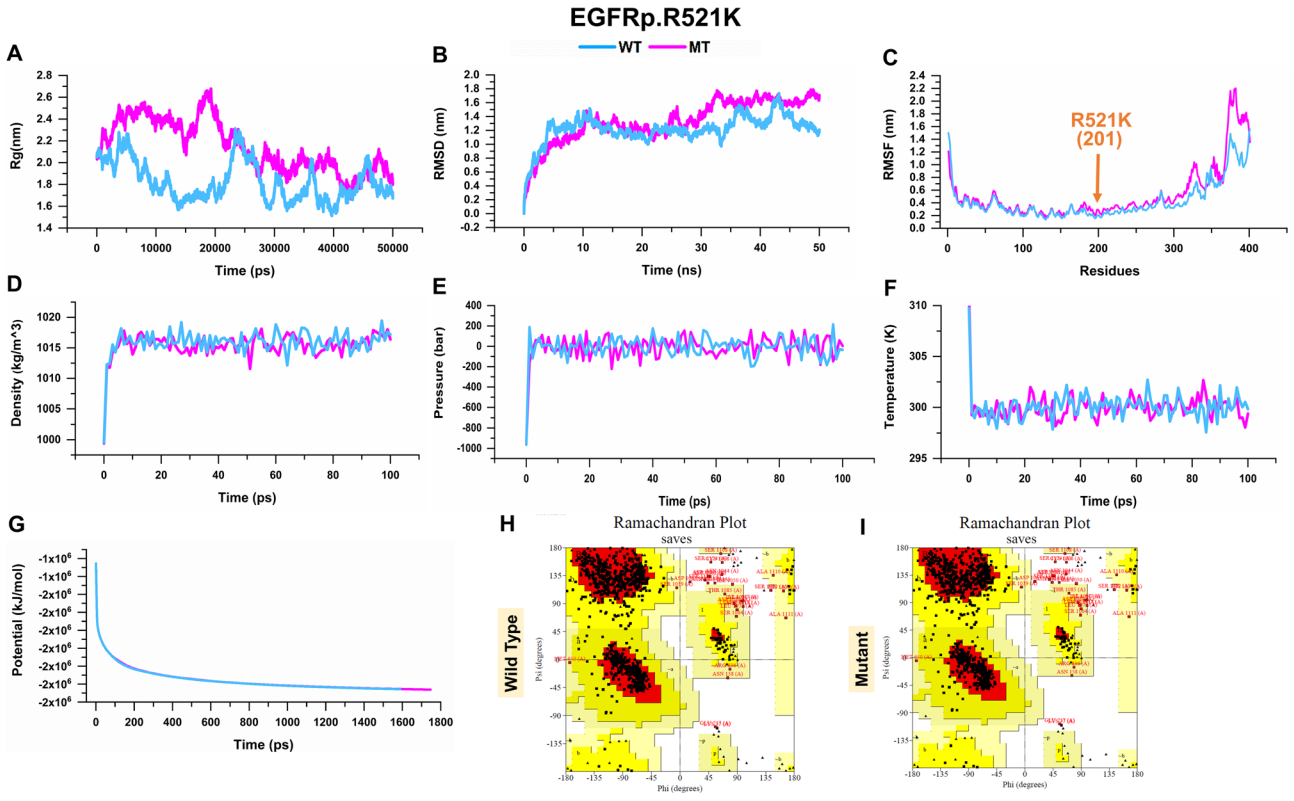


Fig. 7 Analysis of Wild and Mutant *EGFR*$^{p.R521K}$ Using Gromacs Molecular Dynamics Simulations (**A**-**G**) and Ramachandran Plot Profiles (**H**-**I**)

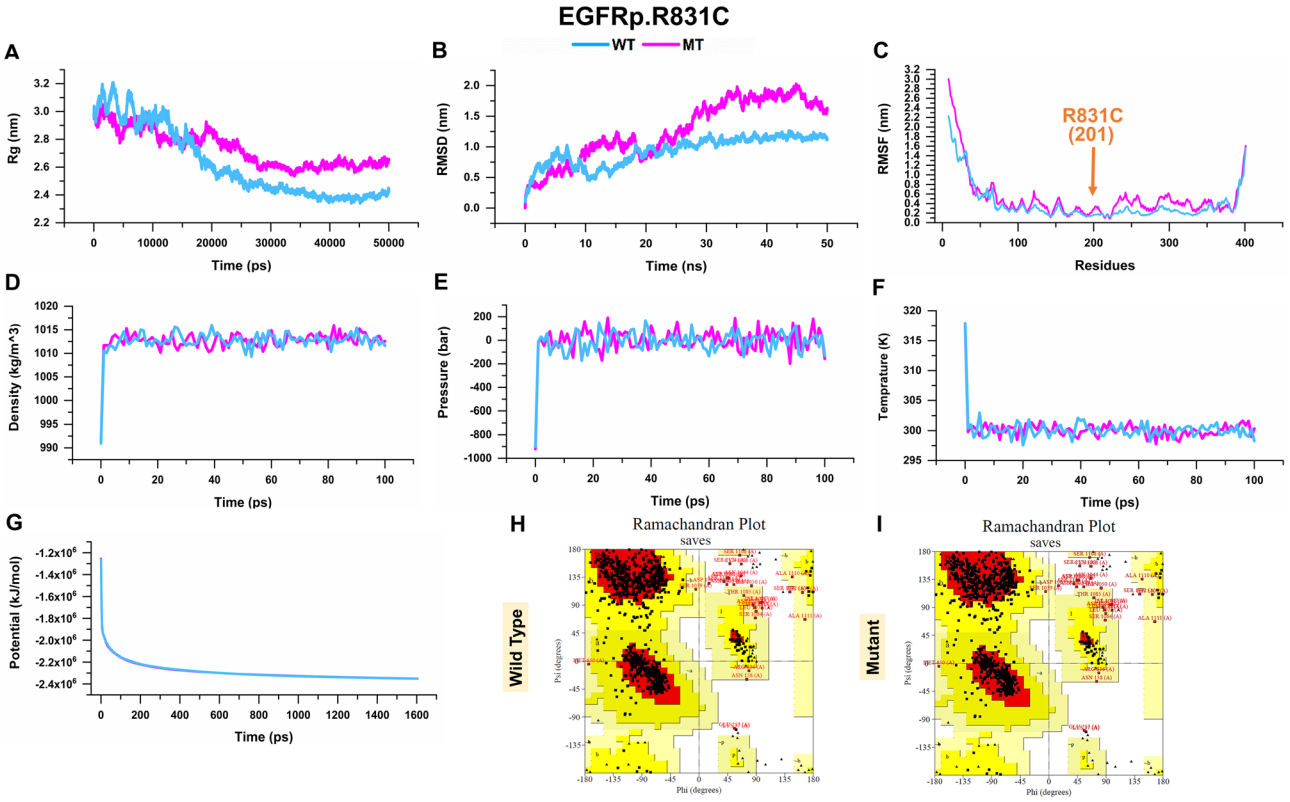


Fig. 8 Analysis of Wild and Mutant *EGFR*$^{p.R831C}$ Using Gromacs Molecular Dynamics Simulations (**A**-**G**) and Ramachandran Plot Profiles (**H**-**I**)

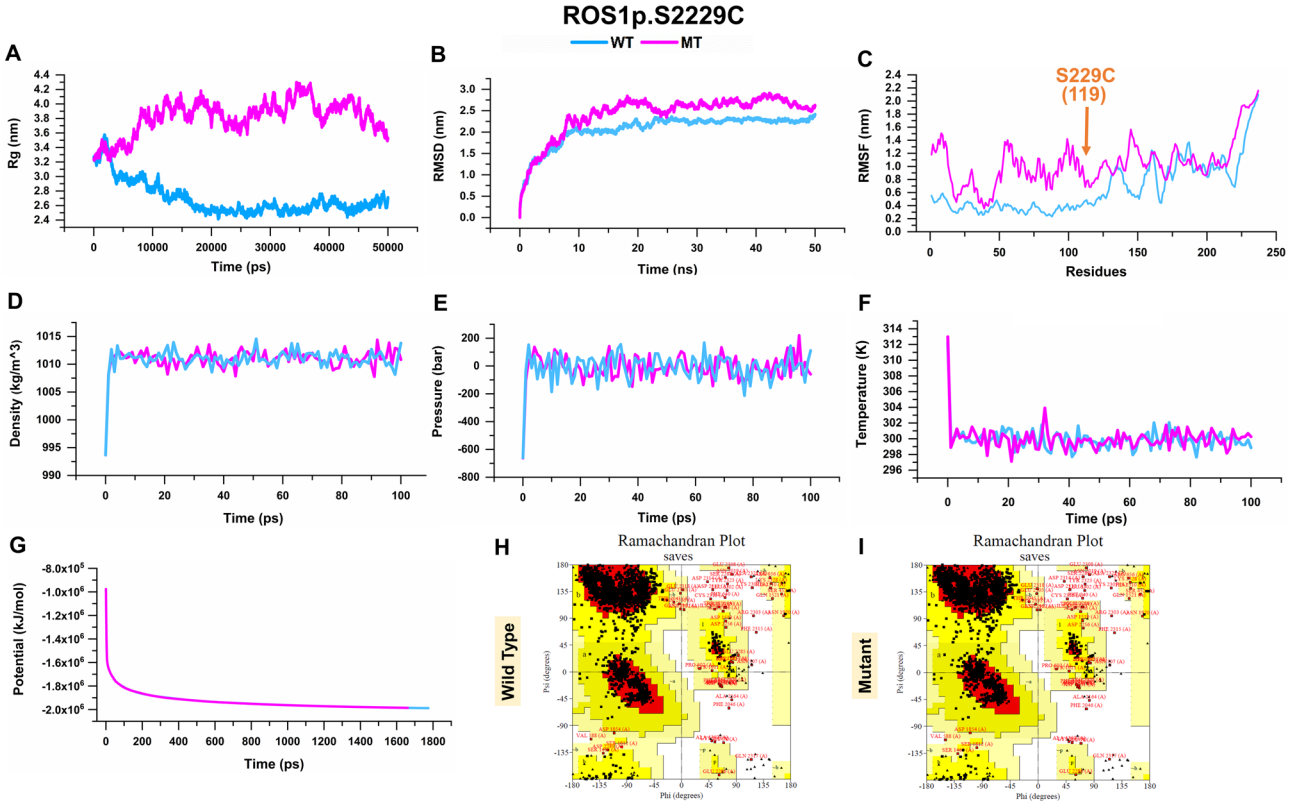


Fig. 9 Analysis of Wild and Mutant *ROS1*$^{p.S2229C}$ Using Gromacs Molecular Dynamics Simulations (**A**-**G**) and Ramachandran Plot Profiles (**H**-**I**)

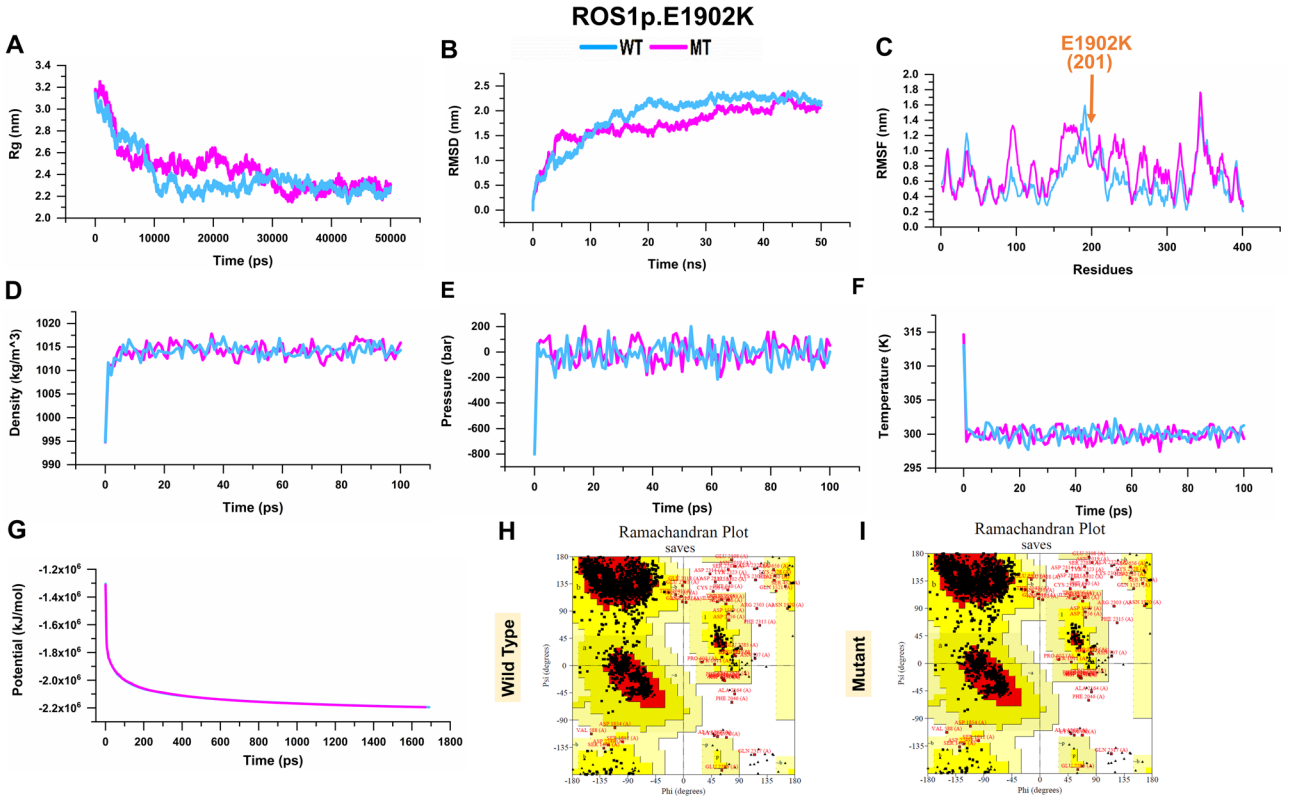


Fig. 10 Analysis of Wild and Mutant *ROS1*$^{p.E1902K}$ Using Gromacs Molecular Dynamics Simulations (**A**-**G**) and Ramachandran Plot Profiles (**H**-**I**)

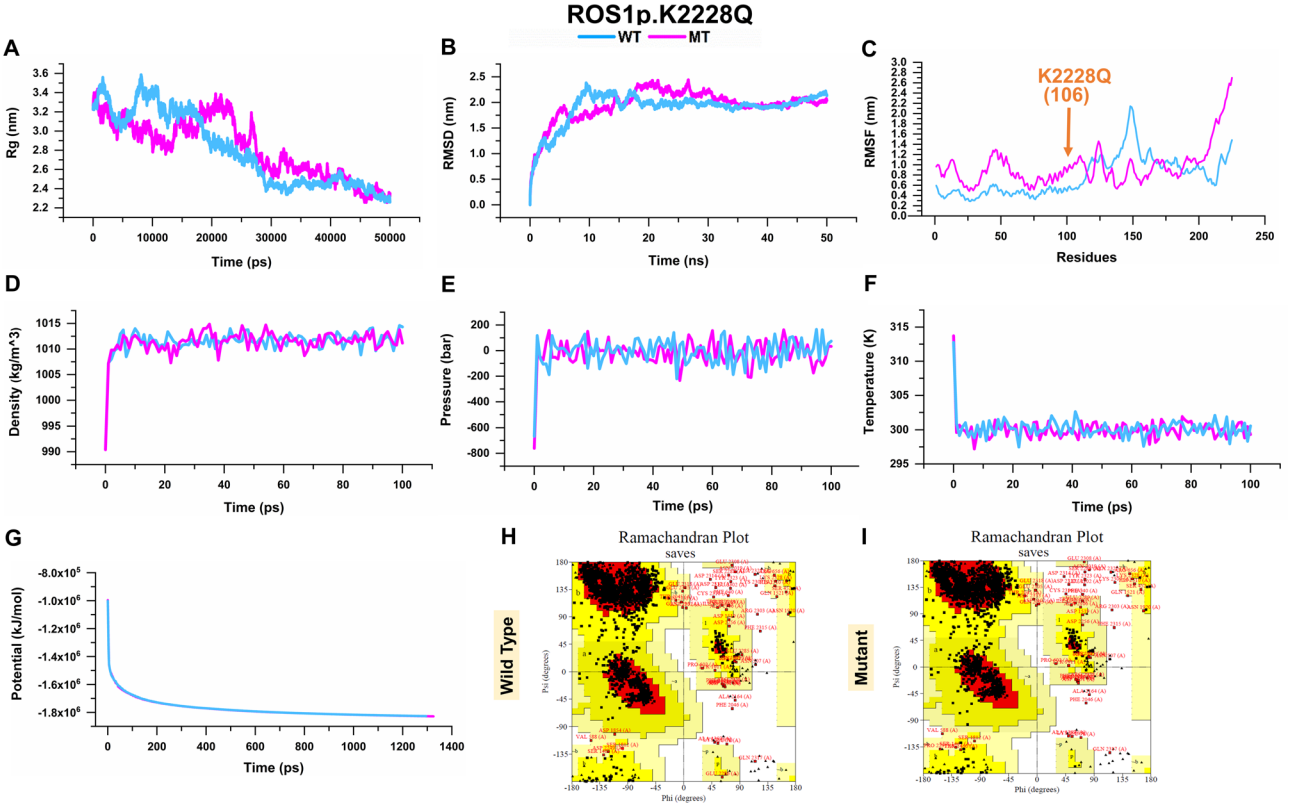


Fig. 11 Analysis of Wild and Mutant *ROS1*$^{p.K2228Q}$ Using Gromacs Molecular Dynamics Simulations (**A**-**G**) and Ramachandran Plot Profiles (**H**-**I**)

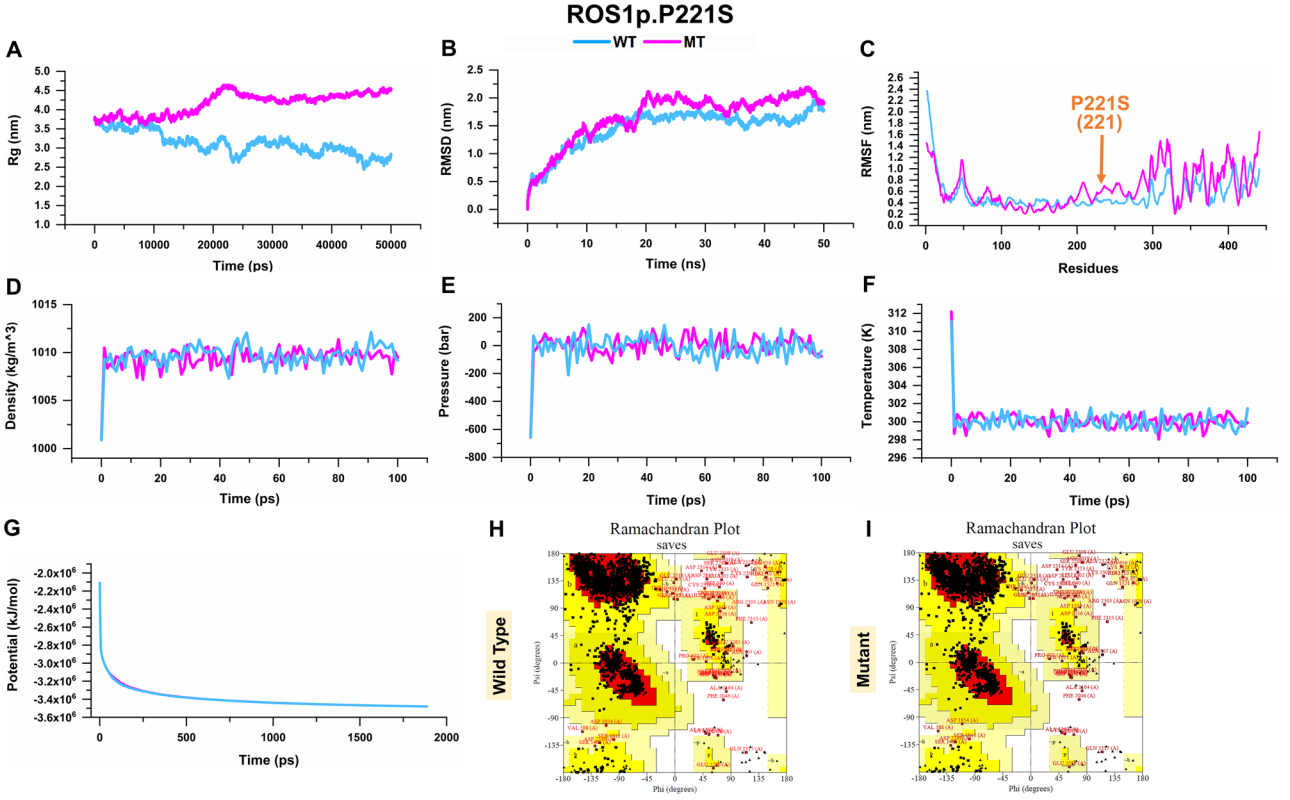


Fig. 12 Analysis of Wild and Mutant *ROS1*$^{p.P221S}$ Using Gromacs Molecular Dynamics Simulations (**A**-**G**) and Ramachandran Plot Profiles (**H**-**I**)

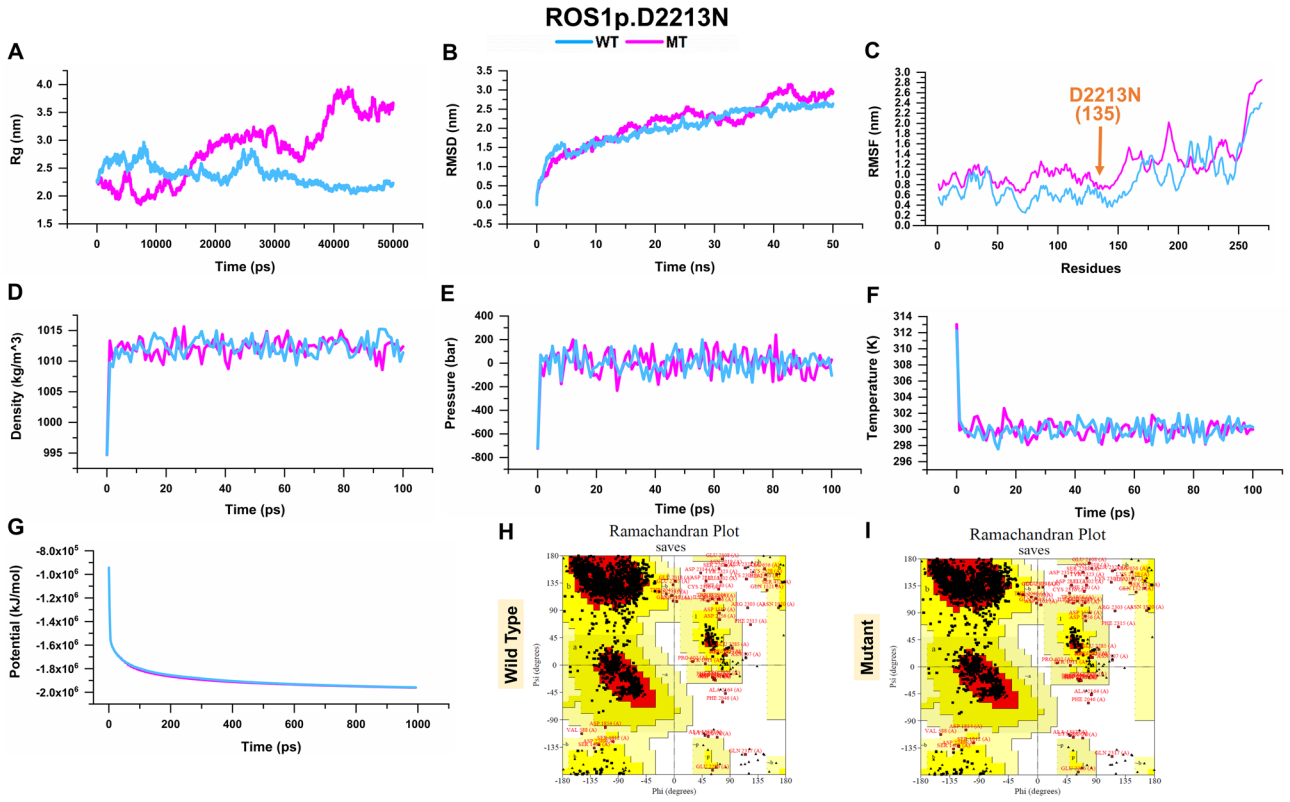


Fig. 13 Analysis of Wild and Mutant *ROS1*$^{p.D2213N}$ Using Gromacs Molecular Dynamics Simulations (**A**-**G**) and Ramachandran Plot Profiles (**H**-**I**)

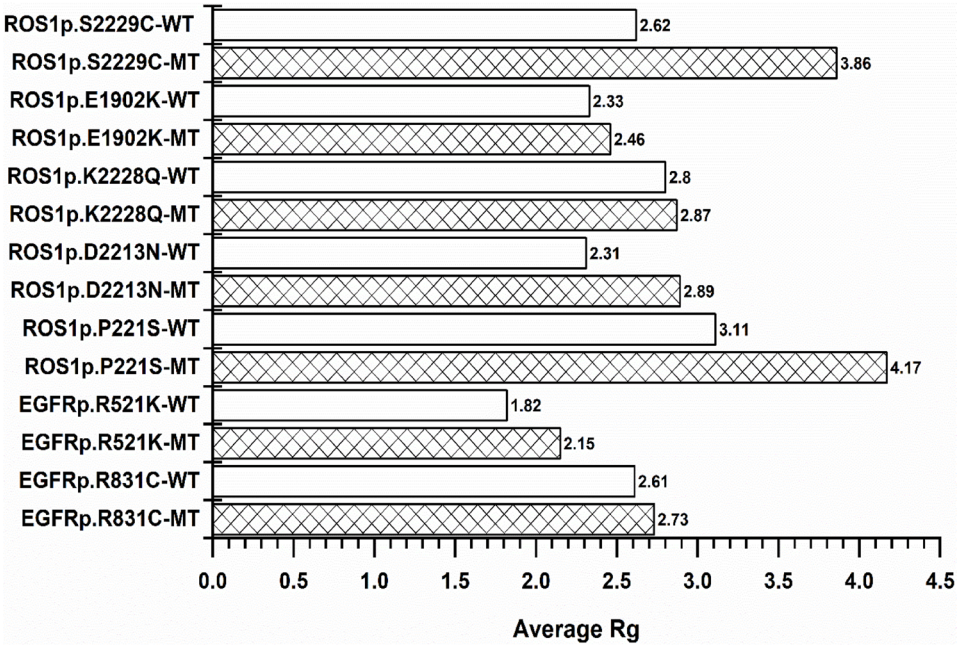


Fig. 14 Average Rg values of interacting sites mutations for *EGFR* and *ROS1* (Both WT and MT)

value; however, the MT was observed to have a more pronounced deviation around the mutation residue. On the other hand, compared with that, in the mutant ($EGFR^{p.R521K}$), fluctuations remain moderate up to residue 200; however, after the mutation site (residue 201 in the graph), the fluctuations become more pronounced, with significant fluctuations observed between residues 300 to 400, as shown in Fig. 7C. Additional energy parameters such as temperature, pressure, and density also exhibited both minor and major fluctuations, demonstrating that these mutations influenced protein stability.

Ramachandran plots were further created for specific interaction site (IS) alterations to examine changes in the favorable areas of both mutant and normal proteins, as illustrated in Figs. 7, 8, 9, 10, 11, 12 and 13 (H-I). For the $ROS1^{p.S2229C}$ (Fig. 9H-I) and $ROS1^{p.P221S}$ (Fig. 12H-I) mutations, the normal protein had 88.2% of its residues in favorable areas, whereas the mutant had 88.3%.

Association of genetic mutations with histopathological grading and sociodemographic parameters

Various clinical, histopathological and socio-demographic charactersitics of the sample were studied in the context of their association with the mutations as indicated in Fig. 15A-E. The analysis of *EGFR*, *BRAF*, *PIK3CA*, *ALK*, and *ROS1* gene mutations revealed significant associations with various demographic and histopathological parameters, highlighting key risk factors for oral cancer progression. Age, gender, tumor differentiation, anatomical tumor location, and tobacco use emerged as prominent factors influencing the mutation patterns. For *EGFR*, the $EGFR^{p.R521K}$ variant was notably associated with younger patients (≤56 years) occurring slightly higher in males (47.4%) as compared to female (37.5%) and with high prevalence in well differentiated tumors (57.1%). This variant also showed a higher occurrence among non-tobacco users (60%) and individuals with dental problems (52.6%), while the $EGFR^{p.R831C}$ mutation was slightly more common in younger patients (≤56 years) and in females (12.5%). We found the $EGFR^{p.R831C}$ mutation in 13.3% of the naswar users' patients. *BRAF* mutations, specifically $BRAF^{p.P253A}$, $BRAF^{p.T526Lfs31}$ and $BRAF^{p.M187Ifs4}$, were linked to well-differentiated tumors (7.14%). The $BRAF^{p.T526Lfs31}$ mutation was found in males (5.26% each), and mutations like BRAFp.^{P253A} and $BRAF^{p.T526Lfs31}$ were found in naswar users (6.66% each). The *PIK3CA* gene mutations, particularly $PIK3CA^{p.I391M}$, were predominantly found in older male patients with moderately-differentiated tumors, primarily located in buccal mucosa. All *PIK3CA* mutations were found in naswar (snuff or dipping tobacco) users and smokers. For the *ALK* gene, the results indicate that different mutations exhibit distinct associations with various clinical and demographic factors. Mutations such as $ALK^{p.S639R}$ and $ALK^{p.G640Efs*25}$ were found to have an association with naswar use (6.66% each) and family history of cancer (8.33% each). In contrast, another mutation i.e. $ALK^{p.T1012M}$ was found only in younger patients (9.09%) and males (5.26%), primarily having a history of naswar (6.66%). In addition, $ALK^{p.K1491R}$ was identified slightly higher in younger patients (81.8%) as compared to older patients (75%) and also showed a strong association with smokers (100%), and naswar users (80%). Additionally,

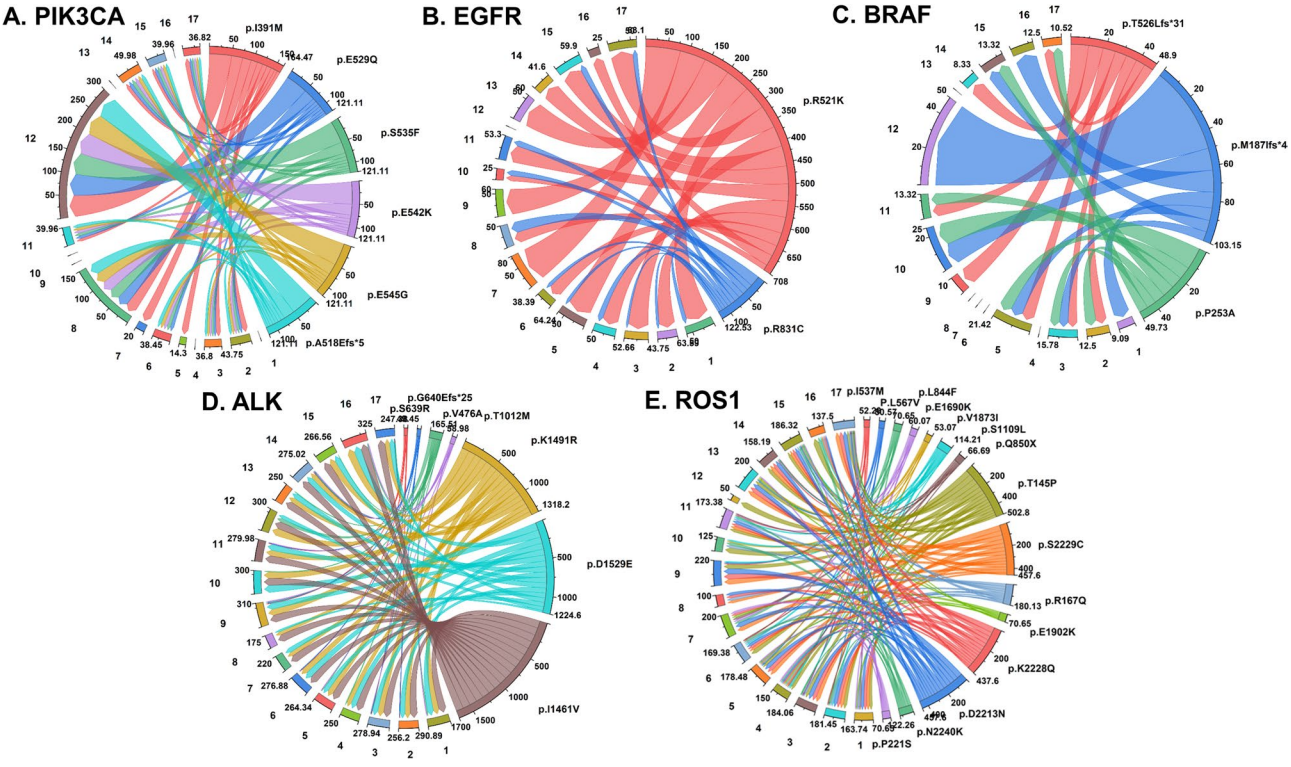


Fig. 15 **A**-**E**. Association of *PIK3CA*, *EGFR*, *BRAF*, *ALK* and *ROS1* genes mutations with histopathological grading and sociodemographic parameters

the *ROS1* mutation p.I537M was more prevalent in males aged < 56 years (9.09%), with a stronger association to moderately differentiated tumors and naswar use. Overall, age, gender, tumor location, and tobacco use, particularly naswar and smoking, emerged as key risk factors influencing the occurrence and distribution of mutations across the analyzed genes. These findings underscore the importance of understanding the interplay between genetic mutations and environmental risk factors in oral cancer progression.

Discussion

Kinase gene mutations drive oncogenesis by disrupting cellular processes like proliferation and apoptosis, making them key targets for OSCC biomarker discovery [20]. Therefore, the hotspot mutations in kinases are considered important for understanding the genetic mechanisms of cancer and tumor biomarker discovery. We present the first detailed mutational analysis of kinase genes in OSCC patients from Khyber Pakhtunkhwa which were further assessed in-silico for pathogenicity and effect on protein using diverse bioinformatics tools. In addition, we performed variant association with diverse demographic and clinicopathological factors. In our study, we uncovered several critical mutations in the five key kinase genes *PIK3CA*, *BRAF*, *EGFR*, *ALK*, and *ROS1*, all of which have established roles in various cancers, including oral squamous cell carcinoma (OSCC).

PIK3CA (Phosphptidylinositol-3-kinase or PI3-kinase) mutations, frequent in various cancers (75%), including OSCC, activate PI3K/AKT signaling, that results in increased cell proliferation and resistance to apoptosis [21, 22]. Our results revealed that out of six mutations, four mutations were located in the adapter domain of *PIK3CA*, i.e., $PIK3CA^{E545G}$, $PIK3CA^{E542K}$, $PIK3CA^{E529Q}$, and $PIK3CA^{S535F}$. In the c2 domain of the *PIK3CA*, one of the pathogenic mutation was identified, i.e., $PIK3CA^{I391M}$. One novel mutation i.e. $PIK3CA^{A518Efs*5}$ was found without localized on any specific domain. A large volume of literature indicates two hotspot mutations ($PIK3CA^{p.E542K}$ and $PIK3CA^{p.E545K}$ at exon 9), in the *PIK3CA* gene, that are associated with nearly 80% of cancers and are considered the charge reversal mutations leading to oncogenic transformation [23, 24]. Interestingly, our findings also revealed the same mutations $PIK3CA^{p.E542K}$ and $PIK3CA^{p.E545G}$ but were positioned on exon 10. Other studies also discovered the hotspot mutations of *PIK3CA* (*E542K* & *E545K)* in the enrolled cancer patients. A study by Zhao et al.. (2022) and Beavers et al..(2021) reported *PIK3CA* hotspot mutations in 60.5% of patients, and ($PIK3CA^{p.E542K}$) in primary tumor tissues using droplet digital PCR supporting the conception that these alterations are common and may have clinical implications for targeted therapy [25, 26]. Another research from a South Indian cohort identified *PIK3CA* mutations in 20% of OSCC cases, predominantly affecting exons 9 and 20

(E545K, H1047Y) [21]. Overall, our results are consistent with previous reports in terms of mutation types and locations (PIK3CA E542K/E545G and EGFR exon 18–21 mutations), but also reveal novel mutation.

EGFR, a transmembrane receptor tyrosine kinase, regulates cell proliferation and survival via *MAPK* and *PI3K* pathways, and its mutated, aberrant signaling is linked to tumor aggressiveness and poor prognosis in cancers including OSCC [27, 28]. Our data revealed *EGFR* mutations in approximately 55.55% (14/27) of cases, which while greater than mutation rates reported in some studies (6.7% of the HNSCC patients) [28]. Previous literature shows mutation on *EGFR* as one of the major reason for cancer progression [29, 30]. However, TCGA data other cohort studies reported lower *EGFR* mutation frequencies, ranging from 7 to 15.8% in Japanese, Asian, Korean, and Greek HNSCC cohorts, with an exceptionally high prevalence of 81.39% observed in a southern Indian population [31, 32]. Furthermore, *EGFR* mutations were predominantly found in exons 18–21, regions frequently and were associated with tumor progression and responsiveness to tyrosine kinase inhibitors (TKIs) [28]. In lung adenocarcinoma *EGFR* mutations primarily occur in exon 19, exons 20 and exon 21 with a rare mutation in exon 18 (3%) [33]. The detection of both somatic ($EGFR^{p.R831C}$ at exon 21) and germline *EGFR* mutations ($EGFR^{p.R521K}$ at exon 13) in our cohort further reinforces its potential as a target for personalized therapies.

BRAF, a serine/threonine kinase and key mediator of the RAS/RAF/MEK/ERK signaling pathway that regulates cell division, harbors most activating non-synonymous mutations in exons 11 and 15 of its kinase domain. These mutations result in constitutive pathway activation, promoting uncontrolled cell growth and tumor progression, and have been implicated in various cancers, including melanoma, thyroid, lung, and colorectal cancers [34, 35]. *BRAF* mutations have been implicated to varying degrees in different cancers like melanoma, thyroid, lung and colorectal cancers etc [36]. Although less common in OSCC, *BRAF* mutations were also detected in our cohort, accounting for 11.11% (3/27) of the mutations. Some studies found no mutations in the *BRAF* genes in OSCC [37, 38]. A study by *Weber et al.* reported a 3% *BRAF* mutation rate in pharyngeal and hypopharyngeal HNSCC specimens but found no mutations in oral cavity tumors [38]. Similarly, Davies et al., Al Sheikh Ali et al.., and Shelly et al.. detected no *BRAF* mutations in HNSCC tissues or oral cancer samples across various populations [37, 39]. We found both nonsynonymous and frameshift deletion mutations in *BRAF*, including the germline $BRAF^{p.P253A}$ mutation and two novel somatic frameshift mutations ($BRAF^{p.M187Ifs*4}$ at exon 4 and $BRAF^{p.T526Lfs*31}$ at exon 12). Their presence in our study suggests that BRAF-targeted treatments could be beneficial for a subset of OSCC patients.

ALK (anaplastic lymphoma kinase) and *ROS1* gene mutations (both receptor tyrosine kinases), are implicated in oral cancer. Though less common, but having an oncogenic nature, mutations in these genes can results in constitutive kinase activity, promotes cell proliferation, and cell cycle progression leading to aberrant activation of downstream signaling pathways and cancer progression [40, 41]. Furthermore, both *ROS1* and *ALK* are involved in fusion events that are key drivers of oncogenesis. These alterations have been identified in several solid tumors, including lung cancer [41, 42] but their roles in OSCC remain less well characterized. Intriguingly, *ALK* mutations were more prevalent in our cohort than previously reported in OSCC, with mutations found in 100% of cases (27/27). These included six nonsynonymous SNVs and a novel frameshift deletion, $ALK^{p.G640Efs*25}$ that was not reported before and was neither predicted with any potential pathogenicity across SIFT, PolyPhen-2, and mutation taster. The higher-than-expected frequency of *ALK* mutations suggests that *ALK* may play a previously underappreciated role in OSCC tumorigenesis, potentially through pathways like PI3K/AKT [43]. The recurrence of mutations such as $ALK^{p.I1461V}$ and $ALK^{p.K1491R}$ positions these as potential biomarkers, highlighting the possibility of ALK inhibitors as a novel therapeutic strategy for OSCC, an area that has been underexplored compared to other cancers.

ROS1 (c-ros oncogene 1) gene harbored the highest number of mutations (15/33; 45.45%) and was considered as the most commonly mutated gene in our study according to mutation rate. In *ROS1*, 06 mutations were located in the fibronectin type III domain, two on the tyrosine kinase domain. In contrast to that, in one study on HNSC, only around 5.04% of patients have harbored somatic mutations in *ROS1* gene [44]. Similarly, other study shows that in 9.9% patients ($n = 10$), *ROS1* mutation was the only mutation detected [45]. *ROS1* gene mutations are known to occur in various cancers, including NSCLC (1-2%), gliomas (6-7%), and cholangiocarcinomas (1.1%), and a number of studies have demonstrated the oncogenic potential of *ROS1* fusions (commonly seen in NSCLC) [46, 47]. We also analyzed the interaction networks for *PIK3CA*, *EGFR*, *BRAF*, *ALK*, and *ROS1* using the STRING server and GeneMania (Figures S2-S3). The interactome pathways shows that these proteins are involved in diverse functions related to genome integrity and any mutation in those proteins can destabilize all the interactions and the cellular process leading to the tumor progression.

In-silico structural analysis using PyMOL revealed significant alterations in the structure of *EGFR* and *ROS1* due to mutations, which may disrupt receptor-ligand interactions and receptor activation, as reported

in previous studies [48]. The evolutionary conservation analysis using ConSurf revealed that five *ROS1* mutations (e.g., $ROS1^{p.N2240K}$, $ROS1^{p.L567V)}$ are located in highly conserved regions (scores 8–9), indicating critical roles in kinase function and protein stability. In contrast, EGFR mutations like $EGFR^{p.R521K}$ (score 3) occur in variable regions, suggesting milder functional impacts. These findings align with prior studies, where conserved mutations in oncogenes, such as $BRAF^{p.V600E}$ and $PIK3CA^{p.H1047R}$, disrupt MAPK and PI3K/AKT signaling, driving oncogenesis in cancers like melanoma and HNSCC [28, 49–51]. This underscores the potential of conserved *ROS1* mutations as OSCC biomarkers in the Pakhtun population, warranting further validation.

Building on our in-silico findings, the identified mutations demonstrate significant therapeutic potential. *PIK3CA* mutations (e.g., $PIK3CA^{p.E542K}$, $PIK3CA^{p.E545G}$) and *EGFR* mutations within exons 18–21 (e.g., $EGFR^{p.R831C}$) are known to be targetable with alpelisib and erlotinib, respectively; however, the emergence of resistance often via activation of downstream effectors such as *AKT* and *MAPK* remains a clinical challenge [52, 53]. The $EGFR^{p.R521K}$ polymorphism has also shown variable responses, with favorable outcomes to cetuximab and 5-FU but limited efficacy with gefitinib [54]. *BRAF* mutations, such as $BRAF^{p.P253A}$, may respond to vemurafenib, yet the presence of frameshift mutations (e.g., $BRAF^{p.M187Ifs*4}$, $BRAF^{p.T526Lfs*31}$) could compromise inhibitor efficacy due to disrupted kinase domain integrity [55]. Additionally, the high prevalence of *ALK* (100%) and *ROS1* (45.45%) mutations suggests potential responsiveness to crizotinib, yet novel frameshifts (e.g., $ALK^{p.G640Efs*25}$) may alter protein conformation and impair drug binding [56]. These findings highlight the existing kinase inhibitors' potential, but resistance via pathway reactivation or structural alterations necessitates personalized strategies and further experimental validation and clinical correlation [52, 53, 55, 56]. Furthermore, beyond therapy, these mutations also carry prognostic value; for instance, *PIK3CA* mutations have been linked to poorer survival outcomes in head and neck cancers [57]. Moreover, as reported by other study, compared to those with wild-type *EGFR*, patients with exon 19 deletions and exon 21 point mutations have shown markedly improved progression-free and overall survival with gefitinib treatment [31] highlighting the importance of mutation profiling in guiding therapy.

Finally, we conducted association of the variants with various demographic and clinicopathological factors and distinct patterns were revealed for all variants. Among them, *ROS1* and *ALK* mutations were found mostly common across the naswar (snuff or dipping tobacco) users and smokers. The variation in mutation frequencies observed in our cohort compared to previous studies may be attributed to a combination of genetic, environmental, and regional factors. Genetic susceptibility and population-specific allele frequencies (population specific molecular architecture) play a critical role in shaping somatic mutational landscapes, particularly in ethnically diverse regions such as South Asia. For instance, EGFR mutations have been shown to vary geographically, with higher prevalence reported in East Asian populations (e.g., 15.1% in Korea) compared to Western cohorts [58]. Similarly, pan-cancer studies have demonstrated significant differences in the frequency of *ALK* and *ROS1* alterations between Asian and Caucasian populations, particularly in lung adenocarcinoma, reflecting a likely influence of ethnic background on mutation patterns [59]. Our study population, based in Khyber Pakhtunkhwa (KPK), represents a unique ethnic group with distinct genetic backgrounds and exposure to specific environmental risk factors, such as naswar (a smokeless tobacco product), cigarette smoking, and betel nut use [60–62]. All these carcinogens may contribute to DNA damage and influence the selection of mutations in key oncogenes including *PIK3CA*, *ROS1*. Together, these factors emphasize the importance of considering ethnic and environmental context when interpreting somatic mutation data in oral cancer.

Conclusion

Gaining insights in to the genetic mutations in oral squamous cell carcinomas not only enable a deeper understanding of the mutational drivers of the cancer but also provide data for tailored therapies. The current state of the knowledge on OSCC, especially from Pakistan, is constrained and the present study unravel the genetic paradigm of OSCC patients using WES for the 1st time. Out of the 33 mutations identified across 5 kinase genes, 84.4% were SNVs. Mutations on the *ALK* i.e. $ALK^{p.I1461V}$, $^{p.K1491R}$, $^{p.D1529E}$ were found in majority of the samples and therefore can be considered as a potential biomarker and may have a significant prognostic value. Results from the ConSurf predictions revealed that 17.3% (5/33) mutations were highly conserved. Molecular dynamics (MD) simulations for Interacting sites (IS) SNVs revealed structural deviations, with mutant proteins revealing larger gyration radius with fluctuating RMSD and RMSF further confirming the perturbed nature of the mutant proteins. The association studies revealed that most variants on *ROS1* were common to naswar (snuff or dipping tobacco) users. This work not only adds to the growing body of evidence underscoring the importance of kinases as both biomarkers and therapeutic targets in cancer but also provides the first detailed genetic landscape of OSCC patients from Khyber Pakhtunkhwa. The identification of recurrent pathogenic mutations in these kinase genes has the potential to inform targeted therapeutic strategies, offering new avenues for

personalized treatment in oral cancer, however, studies in large cohorts are highly recommended.

This study has certain limitations such as a small sample size and being conducted in limited centers, which may have contributed to bias into the final results. Similarly, due to small sample size, it is not possible to effectively integrate molecular research and clinical information.

Acknowledgements
We acknowledge the financial assistance offered by the Higher Education Commission of Pakistan. We are also grateful to the International Center for Chemical and Biological Sciences (Hussain Ebrahim Jamal (HEJ) Research Institute of Chemistry and Dr. Panjwani for Molecular Medicine and Drug Research), University of Karachi for their invaluable assistance and collaboration in this study. We thanks to the patients, their families for their significant contributions to this study. We also acknowledge the University of Sheffield Institutional Open Access Fund for covering the article processing charges.

Authors' contributions
A.A., F.N., S.Y.K., and A.T.K., conceptualize the idea. F.N., W.N., H.A., performed the Laboratory work. F.N., M.F., H.A., A.M., and A.T.K., performed the computational work. A.A., S.F., and A.T.K., supervised the students. M.I., and I.A.K., helped in sequencing. M.K., assisted in patient recruitment. F.N., A.T.K., and H.A., wrote the draft manuscript. A.A., S.F., and M.A., provided significant feedback.

Funding
This study was part of a project titled "ICRG-46: Diagnostic, Prognostic, and Predictive Biomarkers for Oral Squamous Cell Carcinoma," conducted in collaboration with the University of Sheffield in the UK. Higher Education Commission (HEC) of Pakistan supported the Pakistani half of the initiative, while the UK component was funded by the British Council.

Data availability
All the whole exome sequence data is submitted to the NCBI under the project accession number PRJNA1189482; Temporary Submission ID: SUB14714458 and all the SRA record can be accessed at the https://www.ncbi.nlm.nih.gov/bioproject? term=PRJNA1189482&cmd=DetailsSearch.

Declarations

Ethics approval and consent to participate
The research study was approved by the research ethics committee of Khyber Medical University vide reference number *Dir/Ethics/KMU/2020/17* Dated *29/01/2020*.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Institute of Basic Medical Sciences, Khyber Medical University, Phase V, Peshawar 25000, Pakistan
[2]Institute of Pathology and Diagnostic Medicine, Khyber Medical University, Peshawar 25000, Pakistan
[3]Department of Pathology, College of Medicine, Qassim University, Buraidah 52211, Saudi Arabia
[4]School of Cancer Sciences, University of Glasgow, Glasgow G1, UK
[5]Department of Pathology, Lady Reading Hospital Medical Teaching Institution (LRH-MTI), Peshawar, Khyber Pakhtunkhwa 25000, Pakistan
[6]Department of Bioinformatics and Biostatistics, State Key Laboratory of Microbial Metabolism, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China
[7]Phelma Grenoble INP, Université Grenoble Alpes, Grenoble 38000, France
[8]Department of oral and maxillofacial surgery, Khyber college of dentistry, Peshawar, Pakistan
[9]Dr. Panjwani Center for Molecular Medicine and Drug Research, International Center for Chemical and Biological Sciences, Jamil-ur-Rahman Center for Genome Research, University of Karachi, Karachi 75270, Pakistan
[10]Unit of Oral and Maxillofacial Pathology, School of Clinical Dentistry, University of Sheffield, Sheffield S10 2TA, UK

Received: 22 April 2025 / Accepted: 30 June 2025
Published online: 18 August 2025

References

1. Gupta B, Johnson NW, Kumar N. Global epidemiology of head and neck cancers: a continuing challenge. Oncology. 2016;91(1):13–23.
2. Sganzerla JT, Krueger GF, Oliveira MCd, Gassen HT, Santos MAd, Celeste RK, Miguens-Junior SA. Q., relationship between anemia and oral cancer: a case-control study. Brazilian Oral Res. 2021;35:e085.
3. Świętek A, Gołąbek K, Hudy D, Gaździcka J, Biernacki K, Miśkiewicz-Orczyk K, Zięba N, Misiołek M, Strzelczyk JK. The potential association between E2F2, MDM2 and p16 protein concentration and selected sociodemographic and clinicopathological characteristics of patients with oral squamous cell carcinoma. Curr Issues Mol Biol. 2023;45(4):3268–78.
4. Sung H, Ferlay J, Siegel RL, 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. Cancer J Clin. 2021;71(3):209–49.
5. Wéber A, Vignat J, Shah R, Morgan E, Laversanne M, Nagy P, Kenessey I, Znaor A. Global burden of bladder cancer mortality in 2020 and 2040 according to GLOBOCAN estimates. World J Urol. 2024;42(1):1–10.
6. Chamoli A, Gosavi AS, Shirwadkar UP, Wangdale KV, Behera SK, Kurrey NK, Kalia K, Mandoli A. Overview of oral cavity squamous cell carcinoma: risk factors, mechanisms, and diagnostics. Oral Oncol. 2021;121: 105451.
7. Rumgay H, Nethan ST, Shah R, Vignat J, Ayo-Yusuf O, Chaturvedi P, Guerra EN, Gupta PC, Gupta R, Liu S. Global burden of oral cancer in 2022 attributable to smokeless tobacco and areca nut consumption: a population attributable fraction analysis. Lancet Oncol. 2024;25(11):1413–23.
8. Dwivedi R, Pandey R, Chandra S, Mehrotra D. Apoptosis and genes involved in oral cancer-a comprehensive review. Oncol Rev. 2020;14(2):472. https://doi.org/10.4081/oncol.2020.472.
9. Raghavi S, Anbarasu K. Unravelling the role of key genes in oral cancer progression: a comprehensive review. Oral Oncol Rep. 2024;10:100384. https://doi.org/10.1016/j.oor.2024.100384.
10. Su S-C, Lin C-W, Liu Y-F, Fan W-L, Chen M-K, Yu C-P, Yang W-E, Su C-W, Chuang C-Y, Li W-H. Exome sequencing of oral squamous cell carcinoma reveals molecular subgroups and novel therapeutic opportunities. Theranostics. 2017;7(5):1088.
11. Capra M, Nuciforo PG, Confalonieri S, Quarto M, Bianchi M, Nebuloni M, Boldorini R, Pallotti F, Viale G, Gishizky ML. Frequent alterations in the expression of serine/threonine kinases in human cancers. Cancer Res. 2006;66(16):8147–54.
12. Cicenas J, Zalyte E, Bairoch A, Gaudet P. Kinases and cancer. MDPI. 2018;10:63.
13. Jain A, Selvi SGA, Arumugam P, Jayaseelan VP. A computational approach to identify the mutations in the genes of the RTK signaling pathway and their possible association with oral squamous cell carcinoma. Middle East J Cancer. 2021;12(1):1–9.
14. Mayakonda A, Lin D-C, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res. 2018;28(11):1747–56.
15. Ahmad H, Ali A, Ali R, Khalil AT, Khan I, Khan MM, Alorini M. Preliminary insights on the mutational spectrum of BRCA1 and BRCA2 genes in Pakhtun ethnicity breast cancer patients from Khyber Pakhtunkhwa (KP), Pakistan. Neoplasia. 2024;51:100989.
16. Mehmood A, Kaushik AC, Wang Q, Li C-D, Wei D-Q. Bringing structural implications and deep learning-based drug identification for KRAS mutants. J Chem Inf Model. 2021;61(2):571–86.
17. Mehmood A, Nawab S, Jin Y, Kaushik AC, Wei D-Q. Mutational impacts on the N and C terminal domains of the MUC5B protein: a transcriptomics and structural biology study. ACS Omega. 2023;8(4):3726–35.
18. Pickering CR, Zhang J, Yoo SY, Bengtsson L, Moorthy S, Neskey DM, Zhao M, Ortega Alves MV, Chang K, Drummond J. Integrative genomic characterization of oral squamous cell carcinoma identifies frequent somatic drivers. Cancer Discov. 2013;3(7):770–81.

19. Liu X, Wu C, Li C, Boerwinkle E. Dbnsfp v3. 0: a one-stop database of functional predictions and annotations for human nonsynonymous and splice-site SNVs. Hum Mutat. 2016;37(3):235–41.
20. Torkamani A, Verkhivker G, Schork NJ. Cancer driver mutations in protein kinase genes. Cancer Lett. 2009;281(2):117–27.
21. Sathiyamoorthy J, Shyam Sundar V, Babu NA, Shanmugham S, Mani G, Chinnaiyan J, Kalyanaraman P, Hari A. Study on *PIK3CA* gene mutations in oral squamous cell carcinoma among South Indian populations. Biomed Pharmacol J. 2018;11(2):1023–30.
22. Ahmad H, Ali A, Ali R, Khalil AT, Khan I, Khan MM, Alorini M. Mutational landscape and in-silico analysis of TP53, PIK3CA, and PTEN in patients with breast cancer from Khyber Pakhtunkhwa. ACS Omega. 2023;8(45):43318–31.
23. Pang B, Cheng S, Sun S-P, An C, Liu Z-Y, Feng X, Liu G-J. Prognostic role of *PIK3CA* mutations and their association with hormone receptor expression in breast cancer: a meta-analysis. Sci Rep. 2014;4(1):6255.
24. Drury S, Detre S, Leary A, Salter J, Reis-Filho J, Barbashina V, Marchio C, Lopez-Knowles E, Ghazoui Z, Habben K. Changes in breast cancer biomarkers in the IGF1R/PI3K pathway in recurrent breast cancer after tamoxifen treatment. Endocr Relat Cancer. 2011;18(5):565.
25. Fang W, Huang Y, Gu W, Gan J, Wang W, Zhang S, Wang K, Zhan J, Yang Y, Huang Y. PI3K-AKT-mTOR pathway alterations in advanced NSCLC patients after progression on EGFR-TKI and clinical response to EGFR-TKI plus everolimus combination therapy. Transl Lung Cancer Res. 2020;9(4):1258.
26. Borkowska EM, Barańska M, Kowalczyk M, Pietruszewska W. Detection of *PIK3CA* gene mutation in head and neck squamous cell carcinoma using droplet digital PCR and RT-qPCR. Biomolecules. 2021;11(6):818.
27. Lee JW, Soung YH, Kim SY, Nam HK, Park WS, Nam SW, Kim MS, Sun DI, Lee YS, Jang JJ. Somatic mutations of EGFR gene in squamous cell carcinoma of the head and neck. Clin Cancer Res. 2005;11(8):2879–82.
28. Kaur G, Phogat D, Manu V. Study of EGFR mutations in head and neck squamous cell carcinomas. Autopsy Case Rep. 2021;11:e2021251.
29. Panvichian R, Tantiwetrueangdet A, Sornmayura P, Leelaudomlipi S. Missense mutations in exons 18–24 of EGFR in hepatocellular carcinoma tissues. Biomed Res Int. 2015;2015(1):171845.
30. Cserepes M, Nelhűbel GA, Meilinger-Dobra M, Herczeg A, Türk D, Hegedűs Z, Svajda L, Rásó E, Ladányi A, Csikó KG. EGFR R521K polymorphism is not a major determinant of clinical cetuximab resistance in head and neck cancer. Cancers (Basel). 2022;14(10):2407.
31. Vatte C, Al Amri AM, Cyrus C, Chathoth S, Acharya S, Hashim TM, Al Ali Z, Alshreadah ST, Alsayyah A, Al-Ali AK. Tyrosine kinase domain mutations of EGFR gene in head and neck squamous cell carcinoma. OncoTargets Ther. 2017:1527–33.
32. Mirza Y, Ali SMA, Awan MS, Idress R, Naeem S, Zahid N, Qadeer U. Overexpression of EGFR in Oral Premalignant Lesions and OSCC and Its Impact on Survival and Recurrence. Oncomedicine 2018;3:28-36. https://doi.org/10.7150/oncm.22614.
33. Errihani H, Inrhaoun H, Boukir A, Kettani F, Gamra L, Mestari A, Jabri L, Bensouda Y, Mrabti H, Elghissassi I. Frequency and type of epidermal growth factor receptor mutations in Moroccan patients with lung adenocarcinoma. J Thorac Oncol. 2013;8(9):1212–4.
34. Koumaki D, Kostakis G, Koumaki V, Papadogeorgakis N, Makris M, Katoulis A, Kamakari S, Koutsodontis G, Perisanidis C, Lambadiari V. Novel mutations of the HRAS gene and absence of hotspot mutations of the BRAF genes in oral squamous cell carcinoma in a Greek population. Oncol Rep. 2012;27(5):1555–60.
35. Bruckman KC, Schönleben F, Qiu W, Woo VL, Su GH. Mutational analyses of the BRAF, KRAS, and PIK3CA genes in oral squamous cell carcinoma. Oral Surg Oral Med Oral Pathol Oral Radiol Endodontology. 2010;110(5):632–7.
36. Hussain MRM, Baig M, Mohamoud HSA, Ulhaq Z, Hoessli DC, Khogeer GS, Al-Sayed RR, Al-Aama JY. BRAF gene: from human cancers to developmental syndromes. Saudi J Biol Sci. 2015;22(4):359–73.
37. Shelly S, Chien MB, Yip B, Kent MS, Theon AP, McCallan JL, London CA. Exon 15 BRAF mutations are uncommon in canine oral malignant melanomas. Mamm Genome. 2005;16:211–7.
38. Weber A, Langhanki L, Sommerer F, Markwarth A, Wittekind C, Tannapfel A. Mutations of the *BRAF* gene in squamous cell carcinoma of the head and neck. Oncogene. 2003;22(30):4757–9.
39. Ali MAS, Gunduz M, Gunduz E, Tamamura R, Beder L, Tominaga S, Onoda T, Yamanaka N, Grenman R, Shimizu K. Lack of B-RAF mutations in head and neck squamous cell carcinoma. Folia Biol (Praha). 2008;54:157–61.
40. Gainor JF, Varghese AM, Ou S-HI, Kabraji S, Awad MM, Katayama R, Pawlak A, Mino-Kenudson M, Yeap BY, Riely GJ. ALK rearrangements are mutually exclusive with mutations in EGFR or KRAS: an analysis of 1,683 patients with non–small cell lung cancer. Clin Cancer Res. 2013;19(15):4273–81.
41. Yu Z-Q, Wang M, Zhou W, Mao M-X, Chen Y-Y, Li N, Peng X-C, Cai J, Cai Z-Q. ROS1-positive non-small cell lung cancer (NSCLC): biology, diagnostics, therapeutics and resistance. J Drug Target. 2022;30(8):845–57.
42. Takeuchi K, Soda M, Togashi Y, Suzuki R, Sakata S, Hatano S, Asaka R, Hamanaka W, Ninomiya H, Uehara H. RET, ROS1 and ALK fusions in lung cancer. Nat Med. 2012;18(3):378–81.
43. Wang L, Chan HH-Y, Fung AYL, Ng Y-K, Ng PK-S, Su Y, Chan JYK, Poon PHY, Yeung TCK, Tsang CM. Abstract LB414: novel ALK/PIK3CA cooperativity in head and neck squamous cell carcinoma (HNSCC) progression. Cancer Res. 2024;84(7Supplement):LB414-414.
44. Yuan Y, Zhang Q, Duan Q, Tan Y, Chen D. 877P ROS1 mutations as potential negative predictor for response of immunotherapy in patient with head and neck cancer. Ann Oncol. 2023;34:S565.
45. Glaser M, Rasokat A, Prang D, Nogova L, Wömpner C, Schmitz J, Bitter E, Terjung I, Eisert A, Fischer R. Clinicopathologic and molecular characteristics of small-scale ROS1-mutant non-small cell lung cancer (NSCLC) patients. Lung Cancer. 2023;184:107344.
46. Li S, Zhang H, Chen T, Zhang X, Shang G. Current treatment and novel insights regarding ROS1-targeted therapy in malignant tumors. Cancer Med 2024;13(8):e7201.
47. Lin JJ, Shaw AT. Recent advances in targeting ROS1 in lung cancer. J Thorac Oncol. 2017;12(11):1611–25.
48. Galdadas I, Carlino L, Ward RA, Hughes SJ, Haider S, Gervasio FL. Structural basis of the effect of activating mutations on the EGF receptor. Elife. 2021;10:e65824.
49. Getz CGANGscB, I. LMSSCLLCKLEGSB, Wang U. o. K. L. J. W. K.; Drummond Jennifer A. 12 Gibbs Richard A. 12 Kakkar Nipun 12 Wheeler David 12 Xi Liu 12, B. C. o. M. M. D. D. H. K. C. R. J. M. D. H. Y. H. W. C. H. C. K.; 14, U. o. C. S. C. B. I. S. J. B. S. N. S. B. C. Y. C., Comprehensive genomic characterization of head and neck squamous cell carcinomas. Nature. 2015;517(7536):576–582.
50. Shaw AT, Ou S-HI, Bang Y-J, Camidge DR, Solomon BJ, Salgia R, Riely GJ, Varella-Garcia M, Shapiro GI, Costa DB. Crizotinib in ROS1-rearranged non–small-cell lung cancer. N Engl J Med. 2014;371(21):1963–71.
51. Chevallier M, Borgeaud M, Addeo A, Friedlaender A. Oncogenic driver mutations in non-small cell lung cancer: past, present and future. World J Clin Oncol. 2021;12(4):217.
52. Mishra R, Patel H, Alanazi S, Kilroy MK, Garrett JT. PI3K inhibitors in cancer: clinical implications and adverse effects. Int J Mol Sci. 2021;22(7): 3464.
53. Westover D, Zugazagoitia J, Cho B, Lovly C, Paz-Ares L. Mechanisms of acquired resistance to first-and second-generation EGFR tyrosine kinase inhibitors. Ann Oncol. 2018;29:i10-9.
54. Wang Y, Zha L, Liao D, Li X. A Meta-Analysis on the relations between EGFR R521K polymorphism and risk of Cancer. Int J Genomics. 2014;2014(1):312102.
55. Davies H, Bignell GR, Cox C, Stephens P, Edkins S, Clegg S, Teague J, Woffendin H, Garnett MJ, Bottomley W. Mutations of the BRAF gene in human cancer. Nature. 2002;417(6892):949–54.
56. Shaw AT, Solomon BJ. Crizotinib in ROS1-rearranged non-small-cell lung cancer. N Engl J Med. 2015;372(7):683–4.
57. Lui VW, Hedberg ML, Li H, Vangara BS, Pendleton K, Zeng Y, Lu Y, Zhang Q, Du Y, Gilbert BR. Frequent mutation of the PI3K pathway in head and neck cancer defines predictive biomarkers. Cancer Discov. 2013;3(7):761–9.
58. Shigematsu H, Lin L, Takahashi T, Nomura M, Suzuki M, Wistuba II, Fong KM, Lee H, Toyooka S, Shimizu N. Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. J Natl Cancer Inst. 2005;97(5):339–46.
59. Knox MC, Aryamanesh N, Marshall LL, Varikatt W, Weerasinghe C, Burke L, Hau E, Nagrial A, Ashworth S, Kamran SC. Genomic and demographic landscape of non-small cell lung cancer within an ethnically-diverse population–The implications for radiation oncology and personalised medicine. Royal Coll Radiol Open. 2025;3:100341.
60. Humans IWG. O. T. E. o. C. R. T., Betel-quid and areca-nut chewing and some areca-nut derived nitrosamines. IARC Monogr Eval Carcinog Risks Hum. 2004;85:1.
61. International Agency for Research on Cancer. W., Global cancer observatory: cancer today. World Health Organization. 2020.https://gco.iarc.fr/today Accessed 19 Jan 2020.

62. Mutational landscape of. gingivo-buccal oral squamous cell carcinoma reveals new recurrently-mutated genes and molecular subgroups. Nat Commun. 2013;4(1):2873.

Publisher's Note