Article:
Naeem, W., Nawab, F., Sarwar, M.T. et al. (10 more authors) (2025) Profiling genetic mutations in the DNA damage repair genes of oral squamous cell carcinoma patients from Pakistan. Scientific Reports, 15 (1). 7896. ISSN 2045-2322

https://doi.org/10.1038/s41598-025-91700-x

scientific reports

Check for updates

OPEN

Profiling genetic mutations in the DNA damage repair genes of oral squamous cell carcinoma patients from Pakistan

Wafa Naeem[1], Fouzia Nawab[1], Muhammad Tahir Sarwar[1], Ali Talha Khalil[2]✉, Dalia Ali Gaber[3,4], Hilal Ahmad[1], Muhammad Fazeel[5], Mohammed Alorini[6], Ishtiaq Ahmad Khan[7], Muhammad Irfan[7], Muslim Khan[8], Syed Ali Khurram[9]✉ & Asif Ali[6,10,11]✉

Herein, we reported mutations in five DNA Damage Repair (DDR) i.e., *TP53*, *ATR*, *ATM*, *CHEK1* and *CHEK2* involved in OSCC using NG-WES and their analysis using bioinformatics tools. Out of 42 identified mutations, 16.7% are reported for the 1st time. A total of 28 nonsynonymous SNVs are identified. *TP53* harbored the highest number of mutations followed by *ATM*, *ATR*, *CHEK1* and *CHEK2*. Nine mutations ($TP53^{p.R43H}$, $TP53^{p.L125Q}$, $TP53^{p.R116Q}$, $TP53^{p.C110Y}$, $TP53^{p.L62F}$, $ATR^{p.H120Y}$, $ATM^{p.P1054R}$, $ATM^{p.D1853V}$, $ATM^{p.T2934N}$) were predicted highly pathogenic. SAAFEQ-SEQ predicted destabilizing effects for all mutations, while ISPRED-SEQ identified 09 IS mutations, 07 on *TP53*, 01 in *ATR* and 01 in *CHEK1* with no IS mutations predicted for *ATM* and *CHEK2*. Among the IS mutations, only SNVs were used in MDS simulations. The gyration radius for all IS SNVs was larger for mutant as compared to the wild type indicating perturbed folding behavior of the mutant proteins. Structural deviations across the carbon back bone were noted by RMSD for mutant and wild type. The *TP53* IS mutations include $TP53^{p.R116Q}$, $TP53^{p.C110Y}$, $TP53^{p.R43H}$, $TP53^{p.E214X}$, $TP53^{p.R210X}$, $TP53^{p.C110Afs*5}$ and $TP53^{p,S108Ffs*23}$ whereas *ATR* and *CHEK1* IS mutations consist of $ATR^{p.M1932T}$ and $CHEK1^{p.E76Kfs*21}$. ConSurf analysis revealed four SNVs with a high conservation score (9) on *TP53* and *ATM*. $TP53^{p.P33R}$ was predominantly associated with moderately differentiated tumors (84.60%), naswar users (86.60%) and positive family history of cancer (91.60%). The $TP53^{p.P33R}$, $ATR^{p.M211T}$ and $CHEK1^{p.I437V}$ mutations were found recurrently in 21/27 (77.7%), 20/27 (74.04%), and 27/27 (100%) patients, suggesting its potential biomarker applications in local screening.

Keywords Oral cancer, Oncogenomics, NGS, Biomarker, Bioinformatics, Pashtun

Oral cavity cancer is known to be the 6th most common malignancy worldwide[1], accounting for 1.9% of cancer-related deaths[2]. Each year, there are ~ 354,864 new reported cases of oral cancer, representing 2% of all cancer cases[2]. According to the Cancer Statistics 2023, only in the United States there are 34,470 registered new cases and 7440 deaths from oral cavity cancer[3]. In countries such as Pakistan, Bangladesh, India, and Sri Lanka, oral cancer ranks as the second most common cancer, both in terms of its prevalence and its impact on prognosis

[1]Institute of Basic Medical Sciences, Khyber Medical University, Phase V, Peshawar 25000, Pakistan. [2]Department of Pathology, Lady Reading Hospital Medical Teaching Institution (LRH-MTI), Peshawar, Khyber Pakhtunkhwa 25000, Pakistan. [3]Medical Biochemistry and Molecular Biology Department, Faculty of Medicine, Helwan University, Cairo, Egypt. [4]College of Medicine, Gulf Medical University, Ajman, UAE. [5]Phelma Grenoble INP, Université Grenoble Alpes, Grenoble, France. [6]Department of Pathology, College of Medicine, Qassim University, Unaizah, Saudi Arabia. [7]Jamil-Ur-Rahman Center for Genome Research, Dr. Panjwani Center for Molecular Medicine and Drug Research, International Center for Chemical and Biological Sciences, University of Karachi, Karachi 75270, Pakistan. [8]Department of Oral and Maxillofacial Surgery, Khyber College of Dentistry, Peshawar, Pakistan. [9]School of Clinical Dentistry, Faulty of Health, University of Sheffield, Sheffield S10 2TA, UK. [10]Institute of Pathology and Diagnostic Medicine, Khyber Medical University, Phase V, Peshawar 25000, Pakistan. [11]School of Cancer Sciences, University of Glasgow, Glasgow G12 8QQ, UK. ✉email: alitalha.khalil@lrh.edu.pk; s.a.khurram@sheffield.ac.uk; draliasif7@gmail.com

by influencing cancer development, progression, and response to treatment[4]. Oral squamous cell carcinoma (OSCC), the most prevalent neoplasm of the oral cavity (including the palate, floor of mouth, and tongue), accounts for over 90% of oral cancer cases[5] and can result in social isolation, emotional distress, and significant functional impairments as a consequence of both the disease and its treatments[6–8]. The global annual incidence of new OSCC cases is increasing, with the highest rates observed in Asia, followed by Western countries. This trend places OSCC among the 10th most prevalent types of cancer worldwide[9].

In Pakistan, lip and oral cancers are the second most prevalent cancer site overall, accounting for 10.9% of cases across both genders, and the most common cancer among males, with 15.9% of new cases. Risk factors for head and neck cancer are multifaceted and include tobacco use, alcohol consumption[10], chewing habits (areca nut, betel quid, gutka, pan masala, and naswar[11]) and viral infections such as human papillomavirus (HPV) and Epstein-Barr virus (EBV)[10]. Additionally, alterations in oncogenes (e.g., *PIK3CA, RAS*) and tumor suppressor genes (e.g., *TP53, CDKN2A, NOTCH1*) contribute to tumor development[12]. Other nonspecific risk factors include poor oral hygiene, cigar and pipe smoking, and occupational hazards like working in the nickel industry[10]. Some studies question the role of family history in oral cancers[13] while several epidemiological studies suggest a possible correlation between familial cancer history and increased risk for OSCC[14,15]. Additionally, low socioeconomic status (SES) is a key predictor of head and neck squamous cell carcinoma (HNSCC), with a higher incidence reported in low-income populations[16–18]. In the local context, these chewing habits are more prevalent among individuals with lower socioeconomic backgrounds and educational levels, and are particularly common among males[19,20]. In our population, the increasing burden of oral squamous cell carcinoma can be attributed primarily to the chewing habits and the widespread impact of low socioeconomic status. These factors, deeply ingrained in our society, play a significant role in the high prevalence of this disease in Pakistan.

The development of OSCC is also influenced by intrinsic factors, including age and genetic predispositions[21]. Genetic predisposition can influence the functionality of the DNA damage repair (DDR) pathways in oral epithelial cells. These pathways include mismatch repair (MMR), base excision repair (BER), nucleotide excision repair (NER), homologous recombination (HR), and non-homologous end joining (NHEJ), with direct reversal repair (DR) and interstrand crosslink (ICL) repair addressing specific lesions[22]. These mechanisms are crucial for maintaining genomic stability[23]. Double-strand breaks (DSBs), the most severe DNA damage, can be repaired by HR, NHEJ, or both[24]. This complex network of genes, detects DNA lesions, halts the cell cycle to allow for repair, and induces programmed cell death if repair is not feasible[25,26]. Mutations in these DDR genes can affect the efficiency of detecting and repairing DNA lesions which leads to genomic instability[27,28], a key driver of tumorigenesis[29,30].

Head and neck cancer (HNC), including oral squamous cell carcinoma (OSCC), is managed by several treatment modalities like surgery, radiotherapy, immunotherapy, and chemotherapy serving as the cornerstone for locally advanced tumors[31]. Cisplatin, the most widely used chemotherapeutic agent induces cell death by forming DNA adducts and arresting the cell cycle[24,32], and is often combined with agents like paclitaxel, docetaxel, 5-FU (TPF regimen), and immunotherapies (e.g., pembrolizumab, nivolumab) for advanced or metastatic stages[33]. However, only about 40% of patients with locally advanced HNC respond to treatment[34], and the 5-year survival rate remains around 50%[35,36]. Resistance to cisplatin arises through increased DNA repair capacity, particularly via nucleotide excision repair (NER), which plays a key role in clearing cisplatin–DNA adducts[24,32,37]. Targeting DDR pathways, such as through the use of DDR inhibitors (e.g., Mitomycin C, Cisplatin, Etoposide), has emerged as a promising strategy to enhance the effectiveness of DNA-damaging treatments and overcome resistance[38,39].

In OSCC, mutations in the *TP53* gene are associated with reduced survival and resistance to chemotherapy and radiotherapy. *TP53* mutations disrupt a DNA damage response network involving the kinases *ATM* and *ATR*, which activate effector kinases *CHEK1* and *CHEK2* to regulate cell cycle checkpoints and coordinate DNA repair. Disruption of this network impairs the cell's ability to repair DNA and control the cell cycle, contributing to both therapeutic resistance and malignancy in OSCC[40]. In the early stages of oral cancer, the levels of DDR markers like *γ-H2AX, RAD51,* and *53BP1* are higher, showing that the tumor is actively repairing its DNA. Additionally, changes in proteins like *TP53* and *BRCA1*, which are involved in DNA repair, suggest the tumor is using these mechanisms to grow and survive. Because these DDR markers are linked to how the cancer develops, they can be used to predict how the cancer will progress and how well a patient might do (prognosis)[41,42]. Thus these DDR markers may serve as independent prognostic biomarkers.

The identification of distinct mutations in individual cancer samples underscores the importance of characterizing the molecular changes specific to each cancer patient to advance personalized therapeutic approaches[43]. For instance, next-generation sequencing (NGS) technologies have significantly transformed cancer genomics research by offering an extensive approach to detecting somatic alterations in cancer genomes[44,45]. This includes the identification of point mutations, insertions, deletions, and other genetic modifications[46].

A comprehensive analysis at the exome level was conducted to understand the occurrence of mutations in key DNA damage repair genes (*TP53, ATM, ATR, CHEK1* and *CHEK2*) and their correlation with clinicopathological factors. We examined whole-exome sequencing (WES) data from 27 Oral squamous cell carcinoma patients and 7 paired adjacent normal tissues. The novelty of the manuscript lies in the 1st time report of the different mutations through whole exome sequencing from the unexplored Pashtun population belonging to Khyber Pakhtunkhwa. In addition, we have used diverse tools to characterize these mutations and elaborate their structural or functional significance.

Material and methods

Sample selection

Inclusion criteria

Both males and females of all ages which were clinically and histopathologically confirmed for OSCC (Stage I-IV) were included in this study.

Exclusion criteria

Patients with tumor recurrence, those treated with alternative therapies (Chemotherapy, Radiotherapy), or those with other cancers were excluded from the study.

Data collection procedure

Tissue biopsies were collected from patients meeting the inclusion criteria from Hayatabad Medical Complex (HMC) and Khyber College of Dentistry (KCD), Peshawar. Written informed consent (English) was obtained from patients or their guardians and the nature of the study, aims and objectives were explained in local languages (Urdu/Pashto). After obtaining the informed consent, their details were recorded on a structured proforma which included their enrollment ID's, demographic details, clinical investigations and history.

OSCC tissue collection and processing

Tissue samples were collected in 10% formalin under sterile conditions and transported to Khyber Medical University (KMU). Following a 20-h fixation period, the samples were subjected to gross examination for documenting parameters such as size and color. Tumor regions were carefully embedded in cassettes and processed for 16 h in a tissue processor to produce formalin-fixed paraffin-embedded (FFPE) blocks. For slide preparation, 5 µm sections were taken on a microtome and transferred to glass slides, deparaffinized in xylene, and stained with hematoxylin and eosin (H&E) using standard protocols. The H&E-stained slides were examined by an expert pathologist to generate a detailed histopathology report for further analysis. A total of 34 FFPE blocks (27 Tumor, 7 Paired Normal) were selected for Whole Exome sequencing.

DNA extraction and whole exome sequencing

Formalin-fixed paraffin-embedded (FFPE) tumor blocks with high tumor cellularity (> 50%) were selected for DNA extraction. To ensure adequate DNA recovery, a core with a minimum diameter of 2.5 mm, or two cores each with a diameter of 1 mm, were obtained from each block for extraction. The genomic DNA was extracted using the manufacture protocol as per QIAamp DNA FFPE Tissue kit (Catalog No. 56404). The quality and quantity of the extracted DNA was evaluated using 1% agarose gel electrophoresis and Qubit® dsDNA HS Assay Kit (Thermo Fisher) on a Qubit® 2.0 Fluorometer to confirm the presence of an adequate amount of DNA. Successfully passed QC DNA samples were selected for whole exome sequencing.

Following the isolation of DNA from the tumor-rich cores of the FFPE blocks, the quality of DNA was assessed using agarose gel electrophoresis. Only DNA samples with an average fragment size greater than 200 bp were selected for library preparation. A total of 200–300 ng of high-quality DNA was used for library preparation.

The size distribution of the prepared libraries was evaluated using 2% agarose gel electrophoresis. Libraries with a concentration exceeding 10 ng/µL were processed for high-throughput paired-end sequencing on the Illumina platform.

DNA libraries were prepared by fragmenting the DNA and ligating paired-end adapters to the fragments. Exonic regions of the genome were captured by hybridizing the prepared libraries with the Illumina DNA Prep with Exome 2.5 Enrichment Kit, in accordance with the manufacturer's protocol. The captured exonic regions were subsequently amplified to ensure adequate material for sequencing.

The enriched libraries were then sequenced using the Illumina NextSeq 500 platform, employing a 150-bp read length and achieving a mean sequencing depth of 100×, facilitating high-throughput sequencing of the targeted exome regions. The resulting libraries underwent cluster amplification after dilution to a final concentration of 2 nM in 10 µL. Following cluster generation, the flow cells were loaded onto the sequencer. Raw sequencing data were provided in FASTQ format for subsequent analysis. Per-exon coverage for all aligned samples was calculated using Bedtools coverage and the GENCODE (version 47) GTF file. The average exon coverage was 67.25x.

Variant annotation

The FASTQC reports were generated to ensure sequencing quality. FASTQ files were aligned to the human reference genome (hg 38) using the Burrows-Wheeler Aligner (BWA), Following alignment, variant calling and post-alignment processing were conducted using best practices from the Genome Analysis Toolkit (GATK) and Picard-tools-1.109. To minimize false positive detection rates, variants with a depth of coverage (DP) less than 20, a genotype quality (GQ) less than 20, and a quality value (QV) less than 50 were excluded.

For variant annotation the VCF files was then uploaded to ANNOVAR resulting in CSV files with complete information of the variants. R Studio was used for filtering of silent single nucleotide variants (SNVs).

Bioinformatics analysis

Pathogenicity of the mutations, was identified using various in silico prediction tools such as SIFT, PolyPhen-2, Mutation Taster, FATHMM and PROVEAN. The impact of SNVs on the stability of the protein was determined using SAAFEQ-SEQ prediction tool. ISPRED-SEQ prediction tool was used to analyze the interaction site mutations. ConSurf tool was used to evaluate evolutionary conservation and determine the other attributes of the mutated residues such as their nature i.e. functional, structural, buried and exposed.

The mutations were mapped using cBioportal (https://www.cbioportal.org/mutation_mapper). The wild and mutated proteins with mutations on interaction site residues were superimposed and visualized in PyMOL after modelling in the Swiss-Model. Ramachandran plot was generated to compare the wild and mutant type proteins using PROCHECK server.

Molecular dynamic simulations

GROMACS 5.1 was used to perform molecular dynamics simulations. The input file includes a pdb format file having 3D structure of protein generated by Swiss Model. A topology file, which contains the force field parameters and molecule types was generated. Solvation was performed using a solvent model like spc216.gro and ions are added to neutralize the system. Energy minimization step was performed to ensure the system starts from a stable state. NVT (constant number of particles, volume, and temperature) and NPT (constant number of particles, pressure, and temperature) simulations was performed. After the Equilibration of system, the production MD simulation is performed to collect data over the desired time period. After the NVT and NPT simulation, Root Mean Square Deviation (RMSD), radius of gyration (Rg), temperature, pressure, density, and potential energy parameters were calculated. The GROMACS results were visualized using ORIGIN PRO 2024b.

Associations study

The genes mutations were associated with demographic and clinicopathological data of OSCC patients. Various pathological parameter such as (Tumor grade, Tumor site) were included. In addition, we also studied the potential association of the mutations with other social and risk factors such as naswar, tobacco etc. Kaplan–Meier curves were generated to represent the survival probability over time for different mutation categories. This analysis was conducted using R Studio, leveraging the *survival* and *survminer* packages to facilitate the evaluation. The log-rank test was employed to compare survival distributions between groups, allowing us to determine whether the presence of specific mutations significantly impacted overall survival (Fig. 1).

MuTarget based analysis

MuTarget is an open-access platform that enables researchers to link genetic mutations with gene expression changes across various human cancer types. In this study, we utilized MuTarget analysis to identify DDR mutant genes associated with the altered expression of other genes in oral squamous cell carcinoma. A *P*-value threshold of < 0.05 was set to define statistical significance.

An overview of the scheme of study is indicated in Fig. 2.

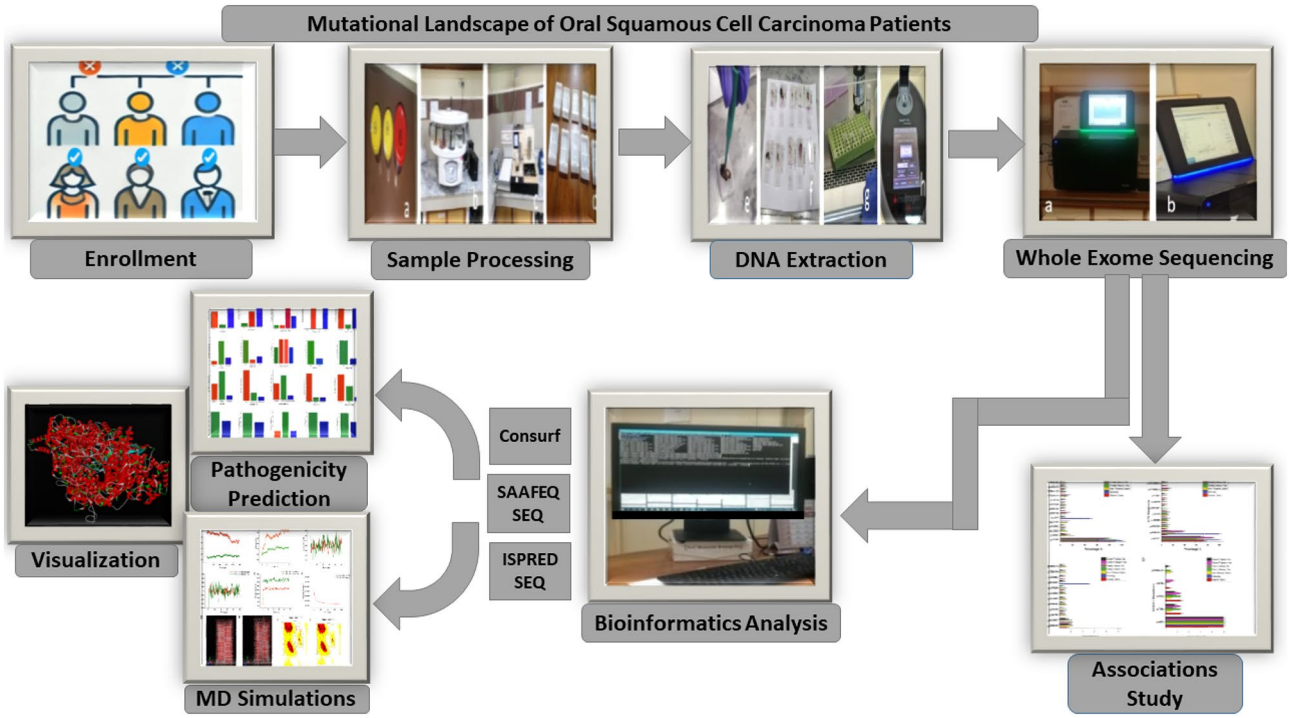

Fig. 1. An Overview of the Mutational Landscape of DNA Damage Repair Genes in Patients with Oral Squamous Cell Carcinoma.

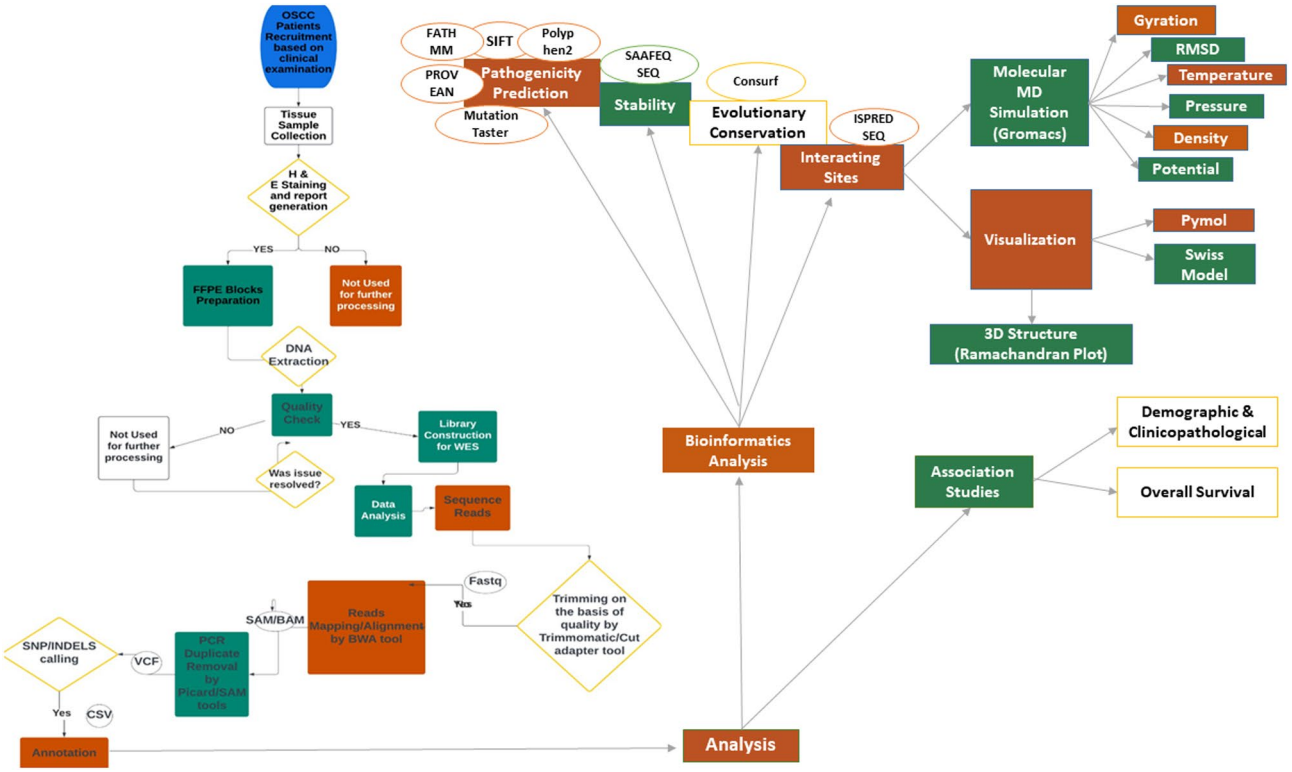

Fig. 2. Scheme of study.

Results

Demographic details

Twenty-seven (27) patients fulfilling the inclusion criteria were enrolled in the study. The study sample included 19 males and 8 females. The cohort included 70.4% of female participants (Fig. 3A) while 59% of the participants aged > 56 years (Fig. 3B). The histopathological Grading revealed that 14/27 (51.8%) were classified as well-differentiated and 13/27 (48.14%) as moderately differentiated. The distribution of OSCC by anatomical site was as follows: 10/27 cases in the tongue, 5/27 in the lip, 4/27 in the buccal mucosa, and 8/27 in other locations such as (Mandible, Oral cavity, floor of mouth, palate of mouth) (Fig. 3C). Concerning tobacco use, 10/27 (37%) patients were non-tobacco users, 15/27 (55.5%) used naswar users, and 2/27 (7.4%) were smokers. Regarding family 12/27 (%) patients revealed positive family history for cancer. Additionally, 8/27 (%) patients had a history of dental problems like infections and mouth swelling etc., whereas 19/27 (70.37%) participants apparently didn't show any historical complications (Fig. 3D). The details are summarized in inset Fig. 3A–D.

Mutational profile

Results from the WES data were analyzed to map mutations on five genes i.e. *TP53, ATR, ATM, CHEK1* and *CHEK2* with potential role in the oncogenic transformation and each were characterized using different parameters. The overall summary of the mutations on these genes indicated in Table 1. Variants found exclusively in tumor tissue samples were categorized as somatic mutations, while variants present in both tumor and paired normal tissue samples were classified as germline mutations[47]. We identified that 16.7% (07/42) of the mutations are not reported previously (Fig. 4A). The identified novel mutations on *TP53* are $TP53^{p.C110Afs*5}$, $TP53^{p.D2Ifs*2}$ and $TP53^{p.W14Cfs*25}$. Other novel mutations identified are $ATR^{p.E125Hfs*9}$, $ATM^{p.E2932Rfs*6}$, $CHEK1^{p.E76Kfs*21}$ and $CHEK2^{p.P448Lfs*51}$. On the *TP53* and *ATR* genes, a single germline mutation was reported i.e. $TP53^{p.P33R}$ and $ATR^{p.M211T}$ while no germline mutations were identified on *ATM* and *CHEK2*. The total germline mutations were reported to be 11.9% (5/42). In total, the somatic mutations were 88.09% (37/42) and were distributed as 92.8%, 90% and 40% on *TP53*, *ATR* and *CHEK1* respectively, while no germline mutations were reported on *ATM* and *CHEK2* (Fig. 4B). The majority of these mutations were nonsynonymous SNVs i.e. 66.6% (28/42), followed by frameshift deletions (7/42) i.e. 16.6% and stop gain mutations (5/42) representing 11.9% (Fig. 4C). Notably, the $CHEK1^{p.I437V}$ mutation was reported in all patients (100%) of oral cancers, whereas, $TP53^{p.P33R}$ and $ATR^{p.M211T}$ were reported in 77.8% and 74.07% patients and therefore, these mutations can be a potential candidate for biomarker applications as indicated in Fig. 4D.

The inset Fig. 5A–E illustrates the lollipop plots of mutations revealing the mutations and their location on their respective genes. Figure 5A–E shows the lollipop plots for *TP53, ATR, ATM, CHEK1* and *CHEK2* respectively.

The inset of Fig. 6A–E reveals the exon wise distribution of the mutations which was found to be highest on exon 3 for *TP53* (7/14; 50%), exon 4 for *ATR* (5/10; 50%), exon 1 for *CHEK1* (3/5; 60%).

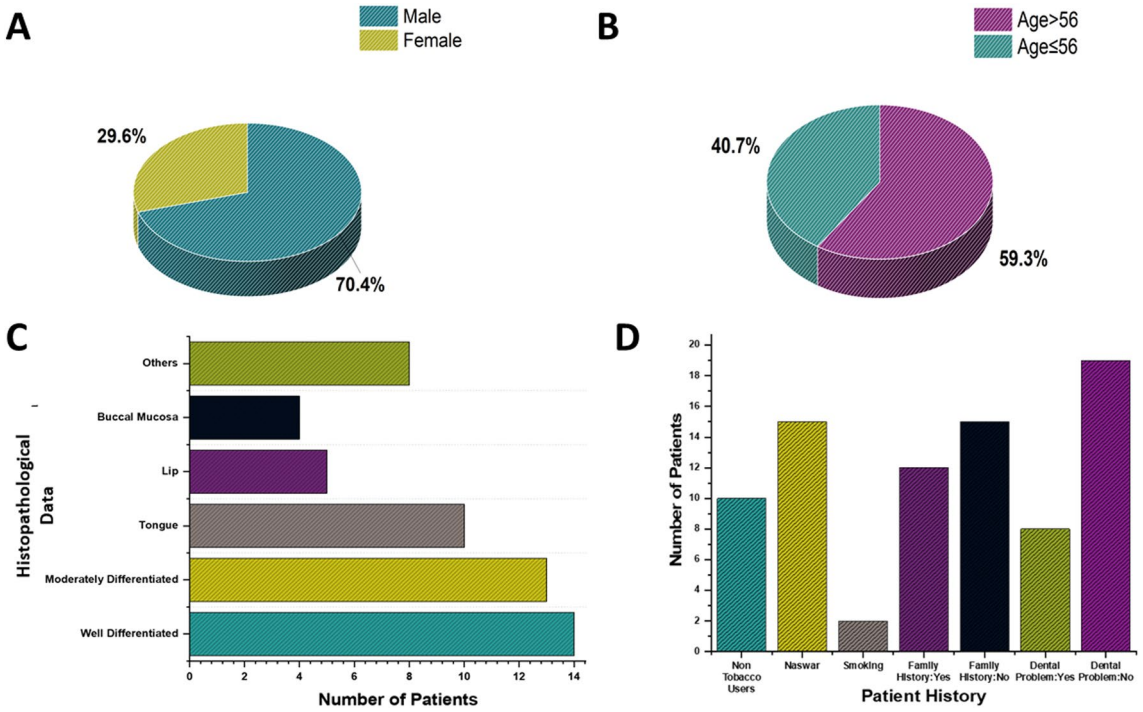

Fig. 3. Sample characteristics and details; (**A**) Gender wise distribution; (**B**) Age wise distribution; (**C**) Histopathological grading and locations; (**D**) General characteristics.

Pathogenicity predictions

SIFT, PolyPhen-2, Mutation Taster, FATHMM and PROVEAN bioinformatics tools were used for the pathogenicity predictions as indicated n Fig. 7A–E.

The predictions from the SIFT database revealed that 10/42 (23.8%) of the mutations were of deleterious nature, whereas, predictions from the PoyPhen-2 database revealed that 7/42 (16.7%) of the mutations were classified under probably damaging. Predictions from Mutation Taster revealed 10/42 (23.8%) mutations as disease causing, whereas, 12/42 (28.5%) were predicted as deleterious by PROVEAN. $TP53^{p.R43H}$, $TP53^{p.L125Q}$, $TP53^{p.R116Q}$, $TP53^{p.C110Y}$, $TP53^{p.L62F}$mutations were predicted as pathogenic across all databases. *ATR* mutations i.e. $ATR^{p.H120Y}$, was reported to be pathogenic by SIFT, PolyPhen-2 and Mutation Taster. *ATM* mutations such as $ATM^{p.P1054R}$ and $ATM^{p.T2934N}$ was predicted as deleterious by SIFT, PolyPhen-2, Mutation Taster and PROVEAN. No predictions are reported for $CHEK2^{p.P448Lfs*51}$. The results are summarized in Fig. 7A–E.

SAAFEQ-SEQ tool was used to predict the effect of mutations on the stability and structural integrity of proteins. All single nucleotide variants were determined to have a destabilizing effect on the respective proteins as indicated in Table S1 Additionally, the interaction site prediction tool (ISPRED-SEQ) was used to identify the interaction site mutations of of *TP53*, *ATR*, *ATM*, *CHEK1* and *CHEK2* as indicated in Table S2. The mutations $TP53^{p.R43H}$, $TP53^{p.R116Q}$, $TP53^{p.C110Y}$, $TP53^{p.E214X}$, $TP53^{p.R210X}$, $TP53^{p.C110Afs*5}$ and $TP53^{p.S108Ffs*23}$ were identified as interaction site (IS) mutations on *TP53*, while $ATR^{p.M1932T}$ and $CHEK1^{p.E76Kfs*21}$ were identified as IS mutations on *ATR* and *CHEK1*, respectively. No IS mutations were detected on *ATM* and *CHEK2*. Overall, 21% of the mutations were predicted as IS mutations. A summary of these ISPRED-SEQ results is provided in Table S2 and Fig. 8.

To determine the evolutionary conservation scores of the mutation residues, ConSurf tool was used and the results are indicated in Fig. 8 and Table S3. ConSurf tool scores the evolutionary conservation based on their structural and functional significance. The results indicate that all *TP53* interaction site mutations ($TP53^{p.R43H}$, $TP53^{p.C110Y}$, $TP53^{p.R116Q}$) and $TP53^{p.L62F}$ are situated in highly conserved regions, each receiving the highest conservation score of 9. $TP53^{p.R43H}$ and $TP53^{p.R116Q}$ were of exposed and functional nature while $TP53^{p.C110Y}$ and $TP53^{p.L62F}$ were predicted as structural residues with buried nature. For *ATM* mutations, $ATM^{p.T2934N}$ was at highest conservation score (9) and predicted as buried while $ATM^{p.P1054R}$ and $ATM^{p.D1853N}$ conservation scores were 8 and were of exposed and functional nature. For *ATR*, the IS mutation $ATR^{p.M1932T}$ was found to be in an exposed region with a conservation score of 4.

PyMOL was used for the superimposition and visualization of the IS mutations as indicated in Fig. 9.

MDS simulations

Based on our ISPRED findings, we shortlisted only the interacting site SNV mutations for molecular dynamic simulations i.e. $TP53^{p.R43H}$, $TP53^{p.R116Q}$, $TP53^{p.C110Y}$ and $ATR^{p.M1932T}$ using GROMACS package 5.1 and various parameters such as Rg (Radius of gyration), root-mean-square deviation (RMSD), pressure, temperature, density, and potential for the mutant and wild type proteins were studied. Ramachandran plots were also generated. The

Gene	Patients	Mutation Type	Reported	Status	Exon	Nucleotide change	Protein change	ClinVar	Alphamissense	SIFT	POLY PHEN 2	MUTATION TASTER	FATHMM	PROVEAN
TP53	21	Nonsynonymous SNV	COSMIC/dbSNP	Germline	3	c.C98G	p.P33R	Likely Benign	–	T	B	P	D	N
	2	Nonsynonymous SNV	COSMIC/dbSNP	Somatic	1	c.G128A	p.R43H	Conflicting Likely Benign/ Uncertain	–	D	D	A	D	D
	1	Nonsynonymous SNV	COSMIC/dbSNP	Somatic	3	c.T374A	p.L125Q	Conflicting Pathogenic/Uncertain	–	D	D	A	D	D
	1	Nonsynonymous SNV	COSMIC/dbSNP	Somatic	3	c.G347A	p.R116Q	Pathogenic/Oncogenic	–	D	D	A	D	D
	1	Nonsynonymous SNV	COSMIC/dbSNP	Somatic	3	c.G329A	p.C110Y	Pathogenic/Likely pathogenic	–	D	D	A	D	D
	1	Nonsynonymous SNV	COSMIC/dbSNP	Somatic	2	c.C184T	p.L62F	Pathogenic/Likely pathogenic	–	D	D	D	D	D
	1	Stop gain	COSMIC	Somatic	6	c.G640T	p.E214X	Pathogenic	–	–	–	A	–	–
	1	Stop gain	COSMIC/dbSNP	Somatic	6	c.C628T	p.R210X	Pathogenic	–	–	–	A	–	–
	1	Stop gain	-	Somatic	4	c.G400T	p.G134X	Pathogenic	–	–	–	A	–	–
	1	Stop gain	COSMIC	Somatic	1	c.C93A	p.Y31X	Pathogenic	–	–	–	A	–	–
	1	frameshift deletion	Novel	Somatic	3	c.327delC	p.C110Afs*5	–	–	–	–	–	–	–
	1	frameshift deletion	COSMIC	Somatic	3	c.322_323del	p.S108Ffs*23	–	–	–	–	–	–	–
	1	frameshift deletion	Novel	Somatic	3	c.4_7del	p.D2Ifs*2	–	–	–	–	–	–	–
	1	frameshift insertion	Novel	Somatic	1	c.38_39insCT	p.W14Cfs*25	–	–	–	–	–	–	–
ATR	20	Nonsynonymous SNV	COSMIC	Germline	4	c.T632C	p.M211T	Uncertain significance	Likely benign	T	B	P	T	N
	5	Nonsynonymous SNV	–	Somatic	42	c.G7082A	p.R2361Q	Benign	–	T	B	P	T	N
	2	Nonsynonymous SNV	COSMIC	Somatic	43	c.G7274A	p.R2425Q	Benign	Likely benign	T	B	P	T	N
	2	Nonsynonymous SNV	–	Somatic	13	c.G2683A	p.V895M	Benign	–	T	B	N	T	N
	2	Nonsynonymous SNV	COSMIC	Somatic	4	c.G946A	p.V316I	Benign	Likely benign	T	B	D	T	N
	1	Nonsynonymous SNV	–	Somatic	34	c.T5795C	p.M1932T	Benign	–	T	B	D	T	N
	1	Nonsynonymous SNV	COSMIC/dbSNP	Somatic	4	c.G891C	p.K297N	Benign	Likely benign	T	B	N	T	N
	1	Nonsynonymous SNV	dbSNP	Somatic	4	c.C358T	p.H120Y	–	Likely benign	D	D	D	T	N
	1	frameshift insertion	dbSNP	Somatic	9	c.2128dupA	p.I710Nfs*3	–	–	–	–	–	–	–
	1	frameshift deletion	Novel	Somatic	4	c.373_377del	p.E125Hfs*9	–	–	–	–	–	–	–
ATM	4	Nonsynonymous SNV	COSMIC	Somatic	37	c.G5557A	p.D1853N	Conflicting Benign/Uncertain	–	T	B	P	T	N
	3	Nonsynonymous SNV	dbSNP	Somatic	28	c.C4138T	p.H1380Y	Benign	Likely benign	T	B	N	T	N
	1	Nonsynonymous SNV	–	Somatic	13	c.A2021G	p.H674R	Conflicting Uncertain/Benign	Likely benign	T	B	N	T	N
	1	Nonsynonymous SNV	dbSNP/COSMIC	Somatic	13	c.T2119C	p.S707P	Benign	Likely benign	T	B	N	T	N
	1	Nonsynonymous SNV	COSMIC	Somatic	22	c.C3161G	p.P1054R	Benign	Ambiguous	D	D	D	T	D
	1	Nonsynonymous SNV	–	Somatic	26	c.G3752T	p.C1251F	Uncertain significance	Likely benign	D	B	D	T	N
	1	Nonsynonymous SNV	COSMIC	Somatic	29	c.C4258T	p.L1420F	Benign	Likely benign	T	B	D	T	D
	1	Nonsynonymous SNV	–	Somatic	35	c.T5256G	p.I1752M	Uncertain significance	Likely benign	T	B	N	T	N
	1	Nonsynonymous SNV	dbSNP	Somatic	37	c.A5558T	p.D1853V	Conflicting Uncertain/Benign	–	D	P	D	T	D
	1	Nonsynonymous SNV	–	Somatic	53	c.A7816T	p.I2606L	–	–	T	B	D	D	N
	1	Nonsynonymous SNV	Not in dbSNP	Somatic	61	c.C8801A	p.T2934N	Uncertain significance	–	D	D	D	T	D
	1	frameshift deletion	Novel	Somatic	61	c.8794delG	p.E2932Rfs*6	–	–	–	–	–	–	–

Continued

Gene	Patients	Mutation Type	Reported	Status	Exon	Nucleotide change	Protein change	ClinVar	Alphamissense	SIFT	POLY PHEN 2	MUTATION TASTER	FATHMM	PROVEAN
CHEK1	27	Nonsynonymous SNV	COSMIC/dbSNP	Germline	12	c.A1309G	p.I437V	Benign	Likely benign	T	B	P	T	N
	5	Nonsynonymous SNV	COSMIC/dbSNP	Germline	1	c.A85G	p.T29A	–	–	T	B	P	T	N
	5	Nonsynonymous SNV	COSMIC/dbSNP	Germline	1	c.C91G	p.P31A	–	–	T	B	P	T	N
	1	Stopgain	COSMIC/dbSNP	Somatic	1	c.G236A	p.W79X	–	–	–	–	A	–	–
	1	frameshift deletion	Novel	Somatic	3	c.226delG	p.E76Kfs*21	–	–	–	–	–	–	–
CHEK2	1	frameshift deletion	Novel		14	c.1342delC	p.P448Lfs*51	–	–		–	–	–	–

Table 1. Mutational Spectrum of *TP53, ATR, ATM, CHEK1* and *CHEK2.*

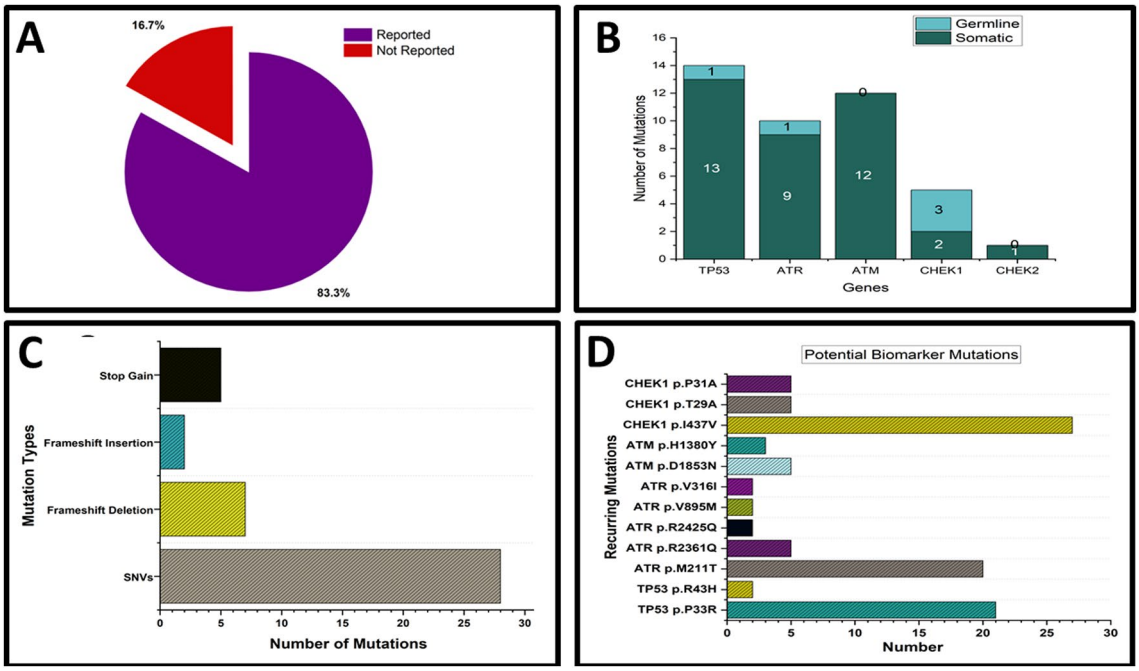

Fig. 4. Characterization of mutations across *TP53, ATR, ATM, CHEK1* and *CHEK2*: (**A**) shows the novel mutations, (**B**) the distribution of germline versus somatic mutations; (**C**) the types of mutations observed within the cohort; and (**D**) mutations with potential biomarker applications.

Fig. 5. Lollipop plots obtained by cBioportal visualization tool Mutation Mapper showing the distribution of different mutations (**A**) *TP53*; (**B**) *ATR*; (**C**) *ATM*; (**D**) *CHEK1* and (**E**) *CHEK2* in enrolled cohort.

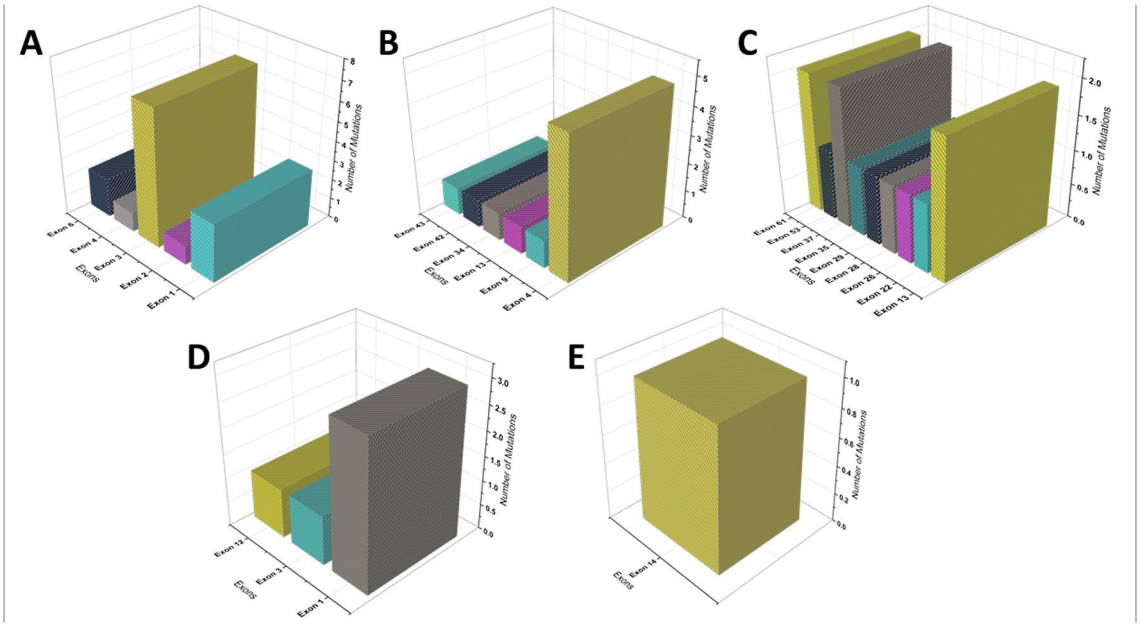

Fig. 6. Frequency of Mutations on Exons: (**A**) *TP53*; (**B**) *ATR*; (**C**) *ATM*; (**D**) *CHEK1* and (**E**) *CHEK2*.

results are indicated in Figure S1A–J to Figure S4A–J for $TP53^{p.R116Q}$, $TP53^{p.R43H}$, $TP53^{p.C110Y}$ and $ATR^{p.M1932T}$ respectively. All the mutant version of the proteins had larger radius of gyration as indicated in Figure S5, which indicates the perturbed protein folding behavior. Generally, a larger Rg value is associated with aberrant protein folding and less stability. Among the studied mutations, the $TP53^{p.R116Q}$, revealed average Rg values of 2.12 nm for wild type and 2.93 nm for the mutant protein. For *ATR* IS mutation i.e. $ATR^{p.M1932T}$, the Rg values for mutant protein was 1.80096 which was higher as compared to the wild type.

Root mean square deviation (RMSD) measurements were used to evaluate the structural changes across the backbone of the mutant and wild type. For $TP53^{p.R43H}$, deviations were detected at 5 ns (Figure S2B). $TP53^{p.C110Y}$ showed a substantial deviation after 4 and 7 ns as indicated in Figure S3B, while $ATR^{p.M1932T}$ exhibited major deviations after 4 ns and minor deviations after 9 ns (Figure S4B). Additional parameters such as temperature, pressure, and density showed both minor and major fluctuations, indicating that these mutations have affected the stability of proteins.

Ramachandran plots for the selected interaction site (IS) mutations were generated to examine changes in the favorable regions of both mutant and normal proteins, as detailed in Figures S1–S4. For the $TP53^{p.C110Y}$ (Figure S3I–J) and $TP53^{p.R43H}$ (Figure S2I–J) mutations, the percentages of residues in favorable regions were found to be 87.7% for the mutant and 88.2% for the normal for each. The $TP53^{p.R116Q}$ showed, with 87.7% and 84.1% of residues in favorable regions for the normal and mutant proteins, respectively (Figure S1I–J). For the $ATR^{p.M1932T}$ mutation, the percentage of residues in favorable regions was 93.3% and 93.2% for normal and mutant proteins (Figure S4I–J).

Clinicopathological associations

Association with demographic and histopathological data

The distribution of various mutations across demographic and clinical factors reveals distinct patterns. The $TP53^{p.P33R}$ mutation is most prevalent among individuals ≤ 56 (9/11; 81.80%) with higher frequencies observed in male participants (15/19; 78.90%). The distribution of $TP53^{p.P33R}$ across tumor site was 70% (7/10) in tongue, 75% (3/4) in buccal mucosa and 80% (4/5) in lip (80%). In addition, $TP53^{p.P33R}$ was found in well differentiated (10/14; 71.4%) and moderate differentiated tumors (11/13; 84.6%) (Figure S6A). The *ATR* mutation, i.e. $ATR^{p.M211T}$ was found prevalent in age group ≤ 56 (9/11; 81.82%), with higher frequencies in males (16/19; 84.21%) and is associated with tumors in the tongue and buccal mucosa, as well as well-differentiated tumors. Mutations like $ATR^{p.V895M}$ and $ATR^{p.V316I}$ were not found in patients in the age group ≤ 56 years and was not found in the female group (Figure S6B).

$ATM^{p.D1853N}$ and $ATM^{p.H1380Y}$ mutations are most common in individuals over 56 years and in females, with $ATM^{p.D1853N}$ showing associations with lip tumor site (3/5; 60%) and moderately differentiated tumors (2/13; 15.38%) (Figure S6C). $CHEK1^{p.I437V}$ mutation is universally present across all ages, genders, and tumor sites, and is associated with both well-differentiated and moderately differentiated tumors. $CHEK1^{p.T29A}$, $CHEK1^{p.P31A}$ and $CHEK1^{p.W79X}$ mutations are more common in older individuals (> 56 years) and associated with moderately differentiated tumors (Figure S6D). The results are summarized in Figure S6A-D.

Association with risk factors

Strong association was observed between $TP53^{p.P33R}$ and naswar users (13/15; 86.60%), with potential relation to family history (11/12; 91.60%). In contrast, variants like $TP53^{p.R43H}$, $TP53^{p.C110Y}$, $TP53^{p.L62F}$, $TP53^{p.E214X}$,

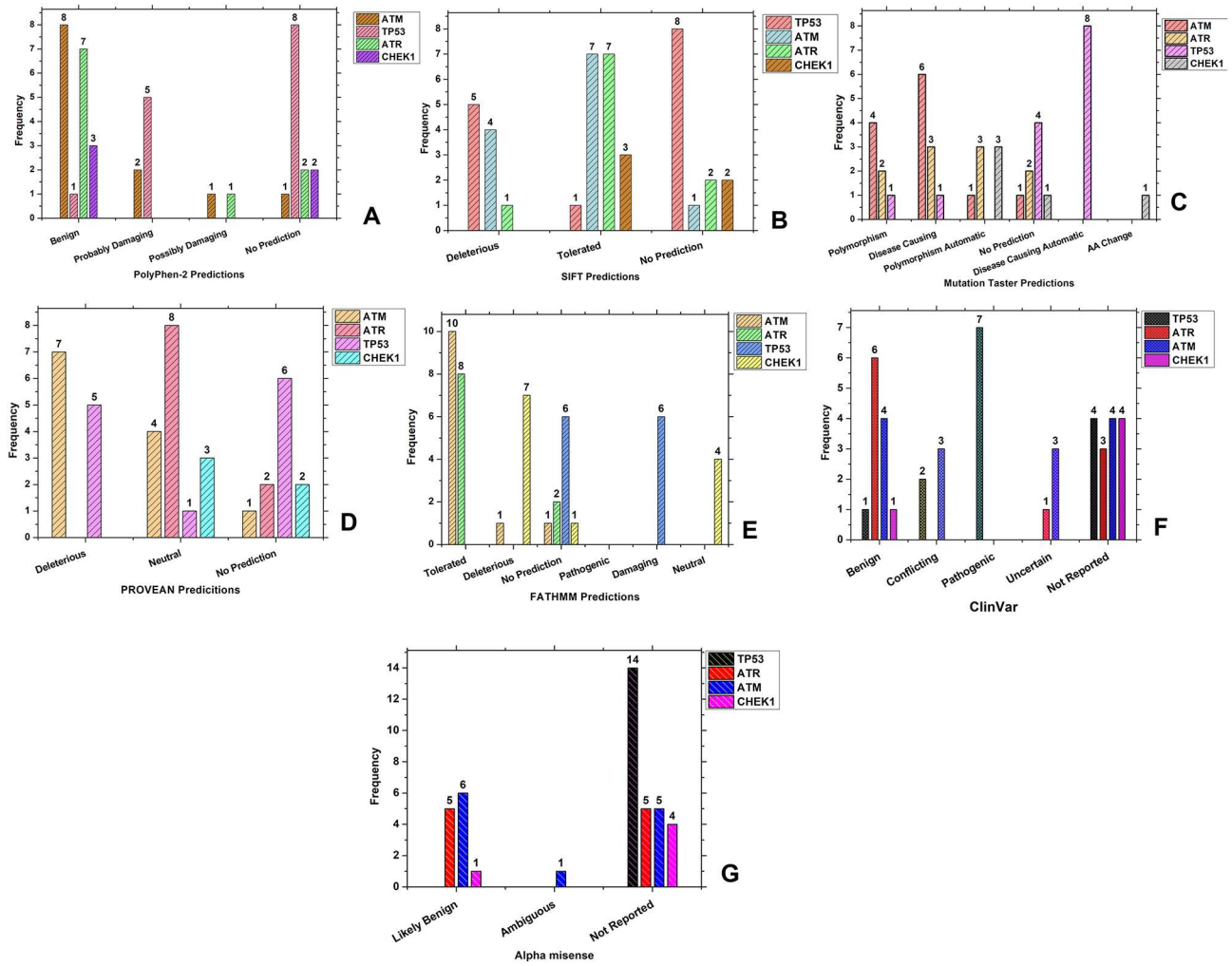

Fig. 7. Pathogenicity of *TP53*, *ATR*, *ATM*, *CHEK1* and *CHEK2* Mutations Based on SIFT, PolyPhen-2, Mutation Taster, FATHMM and PROVEAN: (**A**) SIFT Predictions; (**B**) PolyPhen-2 Predictions; (**C**) Mutation Taster Predictions; (**D**) PROVEAN Predictions and (**E**) FATHMM Predictions, (**F**) ClinVar, (**G**) Alpha missense.

$TP53^{p.C110Afs*5}$ and $TP53^{p.L125Q}$ show no association with naswar users. Variants such as $TP53^{p.R43H}$, $TP53^{p.C110Y}$, $TP53^{p.R116Q}$, $TP53^{p.G134X}$ and all the 3 frameshift mutations were associated with positive family history (Figure S7A). In case of *ATR* mutations $ATR^{p.M211T}$ was found to be prevalent among naswar users (11/15; 73%). Mutations such as $ATR^{p.R2361Q}$ was found only in individuals with a no family history of cancers (Figure S7B). The *ATM* mutations like $ATM^{p.H1380Y}$ was only found in naswar users. Specifically, mutations such as $ATM^{p.D1853N}$ and $ATM^{p.H1380Y}$ are notably prevalent among naswar users and having dental problems, with $ATM^{p.H1380Y}$ also linked to individuals with a positive family history of cancer. Mutations like $ATM^{p.H674R}$, $ATM^{p.C1251F}$, $ATM^{p.L1420F}$, $ATM^{p.T2934N}$, $ATM^{p.E2932Rfs*6}$ and $ATM^{p.P1054R}$ are primarily found in non-tobacco users (Figure S7C). *CHEK1* mutations show that $CHEK1^{p.I437V}$ is present in all groups, while $CHEK1^{p.T29A}$, $CHEK1^{p.W79X}$ and $CHEK1^{p.P31A}$ are common among naswar users and those with a positive family history except $CHEK1^{p.W79X}$. All the mutations except $CHEK1^{p.E76Kfs*21}$ were showed link with dental problems. $CHEK1^{p.E76Kfs*21}$ appears in non-tobacco users (Figure S7D). The results are summarized in Figure S7A–D.

Association with overall survival

We conducted Kaplan–Meier survival analysis using R Studio and its associated libraries, *survival* and *survminer*, to evaluate the association between the identified mutations and overall survival in oral squamous cell carcinoma (OSCC) patients. This analysis aimed to elucidate the impact of specific genetic mutations on patient prognosis.

The results demonstrated that *TP53*, *ATR*, and *ATM* mutations did not exhibit a significant correlation with overall survival, with p-values of 0.64, 0.74, and 0.52, respectively. However, within the *TP53* gene, a significant relationship was observed specifically for patients with the germline $TP53^{p.P33R}$ mutation compared to those with both germline and somatic mutations, with a p-value of 0.012. Patients harboring both germline and somatic mutations in TP53 exhibited a lower overall survival compared to those with only the germline mutation (Fig. 10A–D).

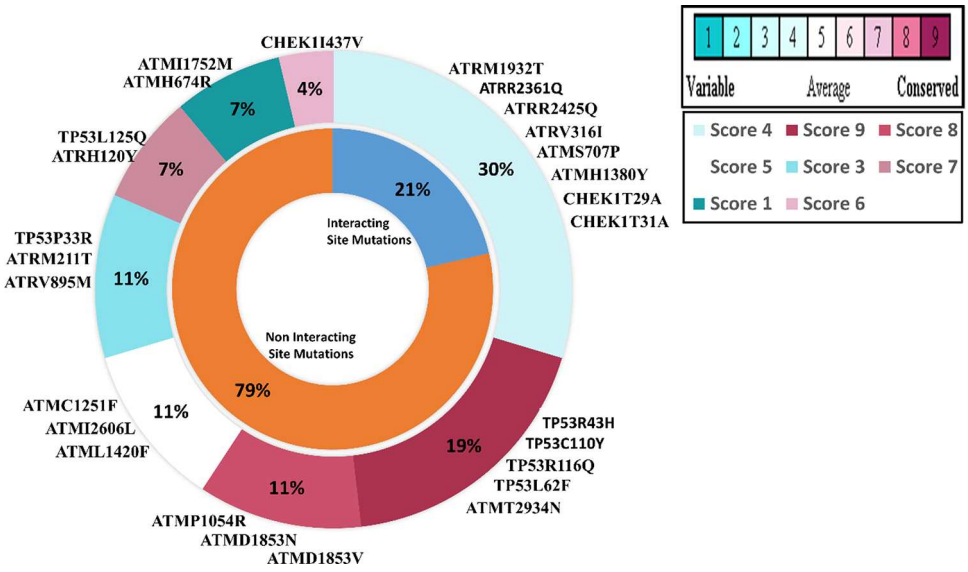

Fig. 8. ISPRED and ConSurf Predictions of *TP53*, *ATR*, *ATM*, *CHEK1* and *CHEK2* Mutations.

MuTarget based analysis of DDR genes

The MuTarget analysis, filtered for cancer hallmark genes, revealed significant findings for *TP53* and *ATR* mutations but showed no associated gene expression changes for *ATM* mutations. In *TP53*-mutant oral squamous cell carcinoma (OSCC), genes such as *SERPINE1*, *CDK6*, *MET*, *MMP10*, and *CAV1* were upregulated, reflecting key cancer hallmarks including cell cycle dysregulation, metastasis, and immune evasion. *SERPINE1* suggests enhanced metastasis, while *CDK6* indicates unchecked cell cycle progression. *MET* promotes tumor motility, and *MMP10* and *CAV1* are linked to increased invasiveness (Figure S8A). In *ATR*-mutant cancers, *FEN1* was downregulated (impairing DNA repair), while *EDIL3* was upregulated, suggesting heightened invasiveness and metastatic potential (Figure S8B). However, for *ATM* mutations, the analysis showed no significant changes in cancer hallmark genes, indicating a lack of associated molecular alterations. Additionally, mutations in *CHEK1* and *CHEK2* also had limited findings due to small sample sizes.

Discussion

The quest for genetic biomarkers in oral squamosus cell carcinoma (OSCC) continues to be a focal point of research, particularly given the role of genetic mutations in disrupting the molecular cascades that regulate DNA damage repair mechanism. The DNA damage repair (DDR) genes and their function is crucial in various aspects of cancer. Deficiencies in DDR mechanisms are well recognized for tumor development and progression[48,49]. Protein interactions, crucial for maintaining normal cellular functions[50], which can be significantly altered by mutations, which results in a truncated protein eventually destabilizing the protein interactome. Using the STRING server and GeneMania, we analyzed the interaction networks for *TP53*, *ATM*, *ATR*, *CHEK1* and *CHEK2* (Figures S9 and S10). STRING analysis identified several key KEGG pathways for *TP53*, including p53 signaling pathway, DNA damage repair (Homologous recombination) and cell cycle checkpoints. For *ATM*, *ATR*, *CHEK1* and *CHEK2*, notable pathways include homologous recombination, fanconi anemia and p53 signaling with exception of mismatch repair pathway exclusively associated with *ATM* (Figure S9). GeneMania interactions further illustrated the extensive signaling networks potentially disrupted by mutations in *TP53*, *ATM*, *ATR*, *CHEK1* and *CHEK2*, contributing to oral squamous cell carcinoma progression as shown in Figure S10.

For pathogenicity analysis 5 bioinformatics tools prediction were utilized. SIFT is a commonly utilized computational tool that predicts the functional effects of mutations by assessing the evolutionary conservation of amino acid residues and their likely impact on protein function[51]. PolyPhen-2 predicts the impact of amino acid changes on protein function by analyzing sequence features and evolutionary conservation, categorizing mutations as benign, possibly damaging, or probably damaging[52]. Mutation Taster predicts the potential pathogenicity of genetic variants by evaluating their likelihood to cause disease based on known mutations and functional data, categorizing them as disease-causing or benign[53]. FATHMM assesses the functional impact of genetic variants on protein function by integrating sequence features and evolutionary conservation, classifying mutations as either damaging or neutral[54,55]. PROVEAN assesses the impact of amino acid substitutions and insertions/deletions on protein function by analyzing sequence conservation and potential functional changes, categorizing mutations as either deleterious or neutral[56].

Research on these key DNA damage repair genes mutations in underrepresented populations, such as the Pakistan, is sparse. *TP53* is a key tumor suppressor that responds to DNA damage by halting the cell cycle at the G1/S or G2/M checkpoints, allowing time for repair. It activates genes involved in DNA repair and, if damage is irreparable, induces apoptosis to prevent tumor development, thereby maintaining genomic stability[28].

Previous literature reports that *TP53* is the most frequently mutated gene across various cancer tissues[57]. Our study also supports this finding, revealing a high frequency of *TP53* mutations in oral squamous cell carcinoma

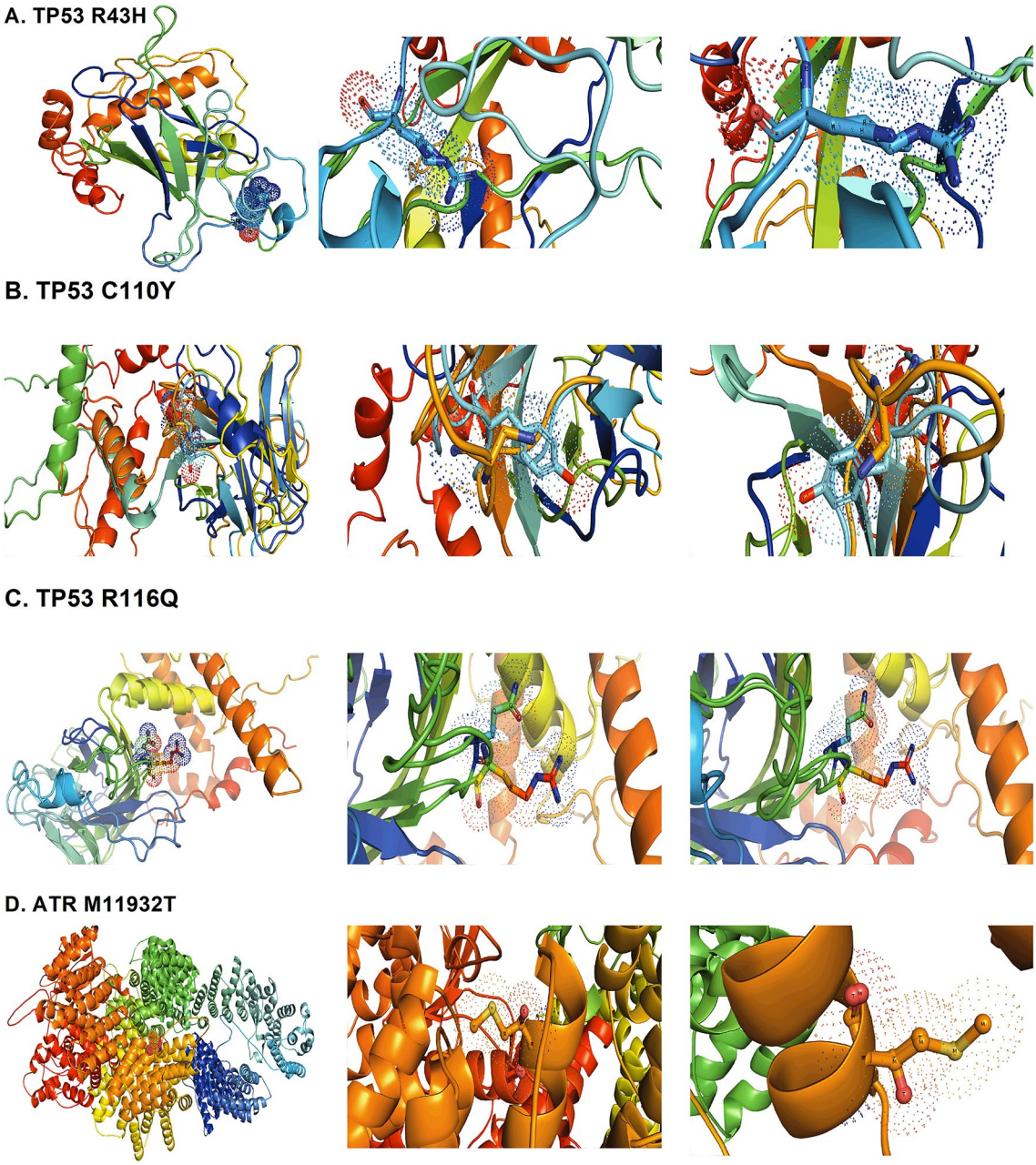

Fig. 9. *TP53* and *ATR* interacting site mutation visualization and superimposition in PyMOL.

patients from Pakistan. In one study *TP53* mutations was observed in 80.6% of OSCC cases and our small cohort reported it in 85.2% cases. The Non synonymous SNVs was reported as the most common mutations in *TP53*, accounting for 64.2% of cases[58]. In our cohort, 6/12 of *TP53* mutations were identified as non-synonymous SNVs, with $TP53^{p.P33R}$ being prevalent that has been reported in lung cancer[59]. Notably, we identified 2 novel frameshift deletion mutation ($TP53^{p.C110Afs*5}$ and $TP53^{p.D2Ifs*2}$) in exon 3. $TP53^{p.C110Afs*5}$ was present on DNA binding domain which has an important role in gene expression. Another frameshift insertion somatic mutation, $TP53^{p.W14Cfs*25}$, was found in the trans activation domain (TAD). The TAD is responsible for the activation of *TP53*. This domain interacts with various transcriptional coactivators and initiates the transcription of *TP53* involved in cell cycle arrest, DNA repair, and apoptosis. Mutations in this domain can impair the protein's ability to activate transcription of its target genes. Overall, the *TP53* mutation sites were diverse but most of the mutations (7/12) were present in DNA binding domain and missense mutations were significantly reported. These OSCC patient findings are similar to previous literature of *TP53* mutations in other cancers[60,61]. Previous studies[62] have shown an association between TP53 mutations and overall survival (OS) ($p = 0.009$). However, in our study, TP53 did not demonstrate a significant correlation with OS ($p = 0.64$). In contrast, when comparing germline mutations to germline + somatic mutations, TP53 showed a significant relationship with OS ($p = 0.012$).

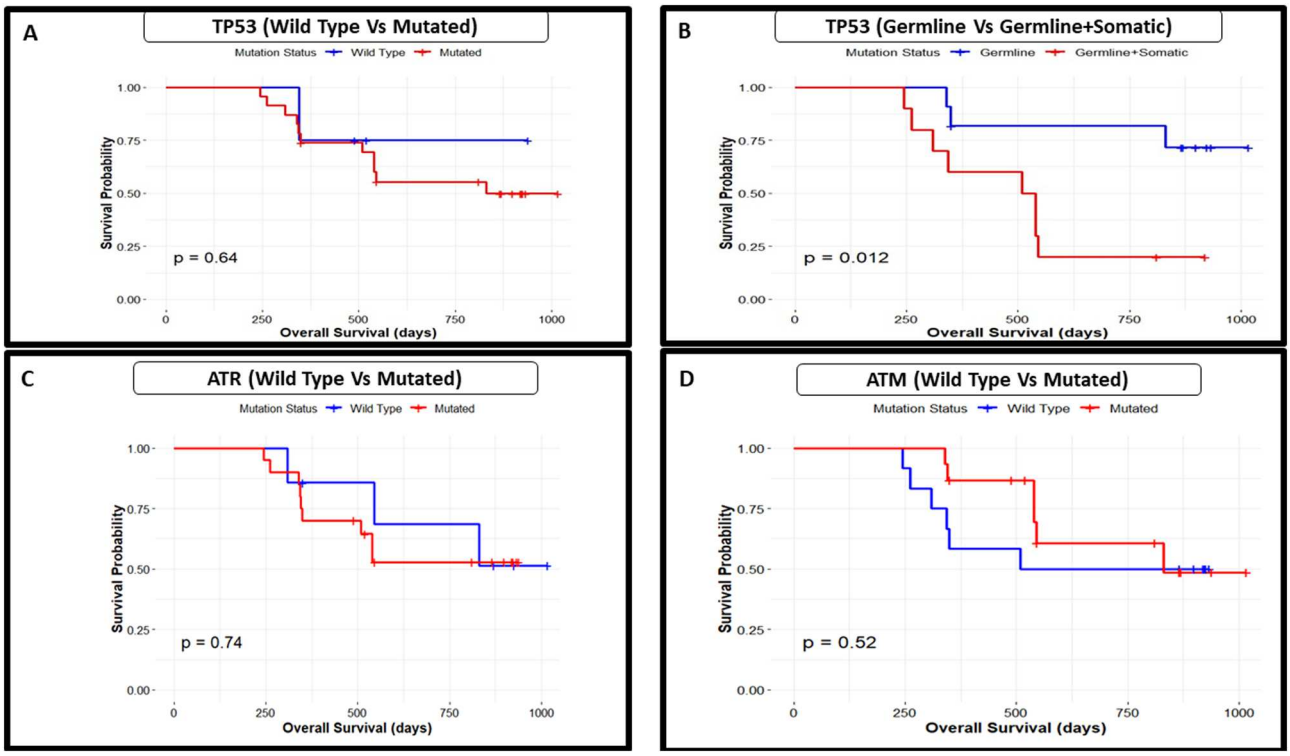

Fig. 10. Kaplan–Meier Plot showing Survival analysis: (**A**) *TP53* Wild Type Vs Mutated; (**B**) *TP53* Germline Vs Germline + Somatic; (**C**) *ATR* Wild Type Vs Mutated; (**D**) *ATM* Wild Type Vs Mutated.

ATM (Ataxia Telangiectasia Mutated) is a crucial protein kinase involved in detecting and responding to DNA damage, particularly double-strand breaks, and coordinating DNA repair and cell cycle regulation[63]. *ATM* mutations are frequently reported in the DNA damage response (DDR) across various cancers[64], particularly in non-small cell lung carcinoma, where *ATM* is identified as the most frequently mutated DDR gene[65] while *ATM* protein expression loss in prior studies has been reported in up to 41% of tumors[66]. Our study supports this literature as *ATM* stands as the 2nd most frequent gene mutated on OSCC samples. It is mainly associated with female sex as reported by R. Biagio et al. in non-small cell lung carcinoma[67]. We identified a novel frameshift mutation ($ATM^{p.E2932Rfs*6}$) in the PI3/PI4 kinase domain, which is critical for regulating cell growth and signaling. This is a truncating mutation and *ATM* truncating mutations can lead to various forms of C-terminally truncated *ATM* proteins. Inherited truncations are associated with ataxia-telangiectasia syndrome, which significantly elevates cancer risk, including a 20% to 30% lifetime risk of lymphoid, gastric, breast, central nervous system, skin, and other cancers[68]. In our cohort of OSCC no germline mutation of *ATM* was reported. *ATM* shows significant relation with OS in metastatic colorectal cancer[69] ($p = 0.01$) while in OSCC cohort its relation with OS was not significant ($p = 0.52$).

ATR (Ataxia Telangiectasia and Rad3 Related) is a crucial protein kinase that plays a central role in managing replicative stress (RS) and regulating the cell cycle. Loss of G1 checkpoint control is nearly universal in cancer, leading to an increased dependence on the S and G2/M checkpoints and *ATR* signaling to manage DNA damage. Thus mutation in *ATR* gene may lead towards cancer progression[70]. In our study one interacting site mutation ($ATR^{p.M1932T}$) of *ATR* was in FAT domain (FRAP, *ATM*, and TRRAP) that facilitates protein–protein interactions and substrate recognition, and 2 mutations ($ATR^{p.R2361Q}$ and $ATR^{p.R2425Q}$) was present in PI3/PI4-kinase domain that is crucial for its kinase activity in phosphorylating substrates involved in DNA damage response and cell cycle regulation. A novel truncating mutation ($ATR^{p.E125Hfs*9}$) was also observed in our cohort. Truncating mutations in *ATR* produce C-terminally truncated proteins, leading to significant genetic instability[71] when combined with mismatch repair deficiencies. In mice, heterozygous *ATR* loss correlates with increased tumorigenesis[72].

CHEK1, a serine/threonine kinase from the CHEK family, plays a crucial role in mediating cell cycle arrest in response to DNA damage. Its action involves activating *ATM* and *ATR*, which phosphorylate *TP53* and other CHEKs, thereby initiating DNA repair mechanism. Its mutations is reported in endometrial, colorectal and stomach carcinomas[73–75]. *CHEK1* is altered in 0.80% of all cancers and mutated in 2.62% of malignant solid tumors[76]. In contrast to this finding our OSCC cohort shows germline mutation of *CHEK1* ($CHEK1^{p.I437V}$) in 100% of samples while other 2 germline mutations ($CHEK1^{p.T29A}$ and $CHEK1^{p.P31A}$) was present in 18.5% of cases. The somatic mutations were recorded in 3.5% of cases which supports the given literature[76].

The *CHEK2* gene encodes checkpoint kinase 2 (CHK2), crucial for the *ATM*-CHK2-p53 pathway that responds to DNA double-strand breaks and prevents early tumorigenesis[77]. Initially linked to moderate breast cancer risk, *CHEK2* mutations are now associated with a wider range of cancers[78,79], making it a key focus in genetic testing for hereditary cancer.

In our study cohort only 1 novel somatic truncating mutation (*CHEK2*$^{p.P448Lfs*51}$) was reported in Pkinase domain which is the central domain responsible for its catalytic activity, structural and regulatory role.

The lack of significant association between these genes and overall survival (OS) may be attributed to the small cohort size. Larger sample sizes are needed to more accurately determine the relationship and potential impact of these genes on OS.

Conclusions

This study analyzed genetic mutations in sporadic oral squamous cell carcinoma (OSCC) patients, focusing on *TP53*, *ATR*, *ATM*, *CHEK1*, and *CHEK2* genes. The cohort of 27 patients revealed a high frequency of somatic mutations, with *TP53* showing the highest mutation frequency. A notable finding was the *CHEK1*p.I437V mutation, present in all patients, suggesting its potential as a biomarker, whereas, *TP53*$^{p.P33R}$ and *ATR*$^{p.M211T}$ was reported in 77.7% and 74% patients indicating their potential as biomarker. Analysis of the mutations' pathogenicity through various bioinformatics tools highlighted the complex nature of predicting their impact, with some mutations, like *TP53*$^{p.R43H}$, consistently predicted as pathogenic. All the 42 mutations were predicted to have a destabilizing effect on protein, which was confirmed through various bioinformatic tools and molecular dynamic simulations. The simulations also showed that the radius of gyration of mutant proteins was higher as compared to the wild type indicating their instability and perturbed folding behavior. The association results showed that the *TP53*$^{p.P33R}$ was found predominantly in the naswar users. Kaplan–Meier survival analysis indicated that while mutations in *TP53*, *ATR*, and *ATM* did not significantly affect overall survival, patients with both germline and somatic *TP53* mutations had a significantly lower survival rate compared to those with only germline mutations. These findings underscore the importance of understanding mutation-specific effects and their potential clinical implications in OSCC.

This study has several limitations, including a small sample size and its conduction across a limited number of centers, which may have introduced bias into the final results. Additionally, due to the small sample size, it was not feasible to effectively coordinate molecular research with clinical data. We recommend further studies in larger cohorts from the same population to further characterize the penetrance of the mutations in OSCC patients. The findings shall help in developing strategies for the management of OSCC patients in local settings.

Data availability

The datasets generated and/or analyzed during the current study are deposited to the NCBI SRA repository; https://www.ncbi.nlm.nih.gov/sra/PRJNA1189482 with scheduled release date on 2025-10-01. The details are; Project accession number PRJNA1189482; Temporary Submission ID: SUB14714458 and Release date: 2025-10-01. Dr. Asif Ali can be contacted at "draliasif7@gmail.com" for any requests related to the data.

Received: 22 October 2024; Accepted: 21 February 2025
Published online: 06 March 2025

References

1. Dhanuthai, K. et al. Oral cancer: A multicenter study. *Med. Oral Patol. Oral Cir. Bucal.* **23**, e23 (2018).
2. Bray, F. et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **68**, 394–424 (2018).
3. Siegel, R. L., Miller, K. D., Wagle, N. S. & Jemal, A. Cancer statistics, 2023. *CA Cancer J. Clin.* **73**(1), 17–48 (2023).
4. Warnakulasuriya, S. J. O. Global epidemiology of oral and oropharyngeal cancer. *Oral Oncol.* **45**, 309–316 (2009).
5. Panarese, I. et al. Oral and Oropharyngeal squamous cell carcinoma: prognostic and predictive parameters in the etiopathogenetic route. *Expert Rev. Anticancer Ther.* **19**, 105–119 (2019).
6. Tenore, G. et al. Tobacco, alcohol and family history of cancer as risk factors of oral squamous cell carcinoma: Case-control retrospective study. *Appl. Sci.* **10**, 3896 (2020).
7. Kunz, V. et al. Screening for distress, related problems and perceived need for psycho-oncological support in head and neck squamous cell carcinoma (HNSCC) patsients: A retrospective cohort study. *BMC Cancer* **21**, 1–10 (2021).
8. Valdez, J. A. & Brennan, M. T. J. D. C. Impact of oral cancer on quality of life. *Dent. Clin.* **62**, 143–154 (2018).
9. Ahmad, W. M. A. W. et al. The predictive model of oral squamous cell survival carcinoma: A methodology of validation. *Biomed Res. Int.* **2021**, 5436894 (2021).
10. Chaudhary, R. K. et al. Identification of hub genes involved in cisplatin resistance in head and neck cancer. *J. Genet. Eng. Biotechnol.* **21**, 9 (2023).
11. Siddiqi, K. et al. Global burden of disease due to smokeless tobacco consumption in adults: Analysis of data from 113 countries. *BMC Med.* **13**, 1–22 (2015).
12. Johnson, D. E. et al. Head and neck squamous cell carcinoma. *Nat. Rev. Dis. Primers* **6**, 92 (2020).
13. Warnakulasuriya, S. J. Causes of oral cancer–an appraisal of controversies. *Br. Dent. J.* **207**, 471–475 (2009).
14. Garavello, W. et al. Family history and the risk of oral and pharyngeal cancer. *Int. J. Cancer* **122**, 1827–1831 (2008).
15. Radoï, L. et al. Family history of cancer, personal history of medical conditions and risk of oral cavity cancer in France: The ICARE study. *BMC Cancer* **13**, 1–10 (2013).
16. Hwang Euna, H. E., Johnson-Obaseki, S., McDonald, J., Connell, C. & Corsten, M. Incidence of head and neck cancer and socioeconomic status in Canada from 1992 to 2007. *Oral Oncol.* **49**, 1072–1076 (2013).
17. Allen, L. et al. Socioeconomic status and non-communicable disease behavioural risk factors in low-income and lower-middle-income countries: A systematic review. *Lancet Glob. Health* **5**, e277–e289 (2017).
18. Johnson, S., McDonald, J. T., Corsten, M. Oral cancer screening and socioeconomic status. *J. Otolaryngol. Head Neck Surg.* **41** (2012).
19. Alam, A. Y. et al. Investigating socio-economic-demographic determinants of tobacco use in Rawalpindi, Pakistan. *BMC Public Health* **8**, 1–9 (2008).
20. Mazahir, S. et al. Socio-demographic correlates of betel, areca and smokeless tobacco use as a high risk behavior for head and neck cancers in a squatter settlement of Karachi, Pakistan. *Subst. Abuse Treat. Prev. Policy* **1**, 1–6 (2006).
21. Du, E. et al. Long-term survival in head and neck cancer: Impact of site, stage, smoking, and human papillomavirus status. *The Laryngoscope* **129**, 2506–2513 (2019).
22. Chatterjee, N. & Walker, G. C. Mechanisms of DNA damage, repair, and mutagenesis. *Environ. Mol. Mutagen.* **58**, 235–263 (2017).

23. Colombo, C. V., Gnugnoli, M., Gobbini, E. & Longhese, M. P. J. How do cells sense DNA lesions?. *Biochem. Soc. Trans.* **48**, 677–691 (2020).
24. Rocha, C. R. R., Silva, M. M., Quinet, A., Cabral-Neto, J. B. & Menck, C. F. M. J. C. DNA repair pathways and cisplatin resistance: An intimate relationship. *Clin.* **73**, e478s (2018).
25. Tubbs, A. & Nussenzweig, A. J. C. Endogenous DNA damage as a source of genomic instability in cancer. *Cell* **168**, 644–656 (2017).
26. Jiang, M. et al. Alterations of DNA damage repair in cancer: From mechanisms to applications. *Ann. Transl. Med.* **8**(24), 1685–1685 (2020).
27. Groelly, F. J., Fawkes, M., Dagg, R. A., Blackford, A. N. & Tarsounas, M. Targeting DNA damage response pathways in cancer. *Nat. Rev. Cancer* **23**, 78–94 (2023).
28. Vaddavalli, P. L. & Schumacher, B. J. The p53 network: Cellular and systemic DNA damage responses in cancer and aging. *Trends Genet.* **38**, 598–612 (2022).
29. Lord, C. J. & Ashworth, A. J. S. PARP inhibitors: Synthetic lethality in the clinic. *Science* **355**, 1152–1158 (2017).
30. Carusillo, A. & Mussolino, C. J. C. DNA damage: From threat to treatment. *Cells* **9**, 1665 (2020).
31. Muzaffar, J., Bari, S., Kirtane, K. & Chung, C. H. J. C. Recent advances and future directions in clinical management of head and neck squamous cell carcinoma. *Cancers* **13**, 338 (2021).
32. Aldossary, S. A. J. B. & Journal, P. Review on pharmacology of cisplatin: Clinical use, toxicity and mechanism of resistance of cisplatin. *Biomed. Pharmacol. J.* **12**, 7–15 (2019).
33. Silva, J. P., Pinto, B., Monteiro, L., Silva, P. M. & Bousbaa, H. J. P. Combination therapy as a promising way to fight oral cancer. *Pharm.* **15**, 1653 (2023).
34. Lo Nigro, C., Denaro, N., Merlotti, A. & Merlano, M. J. Head and neck cancer: improving outcomes with a multidisciplinary approach. *Cancer Manag. Res.* **2017**, 363–371 (2017).
35. Bahadır, A. et al. Protective effects of curcumin and beta-carotene on cisplatin-induced cardiotoxicity: An experimental rat model. *Anatol. J. Cardiol.* **19**, 213 (2018).
36. Mohapatra, P. et al. CMTM6 drives cisplatin resistance by regulating Wnt signaling through the ENO-1/AKT/GSK3β axis. *JCI Insight* **6**(4), e143643 (2021).
37. Zhou, J. et al. The drug-resistance mechanisms of five platinum-based antitumor agents. *Front. Pharmacol.* **11**, 343 (2020).
38. Helleday, T., Petermann, E., Lundin, C., Hodgson, B. & Sharma, R. A. J. DNA repair pathways as targets for cancer therapy. *Nat. Rev. Cancer* **8**, 193–204 (2008).
39. Martin, S. A., Lord, C. J. & Ashworth, A. J. DNA repair deficiency as a therapeutic target in cancer. *Curr. Opin. Genet. Dev.* **18**, 80–86 (2008).
40. Lindemann, A., Takahashi, H., Patel, A., Osman, A. & Myers, J. J. Targeting the DNA damage response in OSCC with TP53 mutations. *J. Dent. Res.* **97**, 635–644 (2018).
41. Nikitakis, N. G. et al. Alterations in the expression of DNA damage response-related molecules in potentially preneoplastic oral epithelial lesions. *Oral Pathol. Oral Radiol.* **125**, 637–649 (2018).
42. Oliveira-Costa, J. P. et al. BRCA1 and γH2AX as independent prognostic markers in oral squamous cell carcinoma. *Oncoscience* **1**, 383 (2014).
43. Ahmad, H. et al. Clinico-genomic findings, molecular docking, and mutational spectrum in an understudied population with breast cancer patients from KP. *Pakistan. Front. Genet.* **15**, 1383284 (2024).
44. Ahmad, H. et al. Mutational landscape and in-silico analysis of TP53, PIK3CA, and PTEN in patients with breast cancer from khyber Pakhtunkhwa. *ACS Omega* **8**, 43318–43331 (2023).
45. Ahmad, H. et al. Preliminary insights on the mutational spectrum of BRCA1 and BRCA2 genes in Pakhtun ethnicity breast cancer patients from Khyber Pakhtunkhwa (KP), Pakistan. *Neoplasia* **51**, 100989 (2024).
46. Nakagaki, T. et al. Profiling cancer-related gene mutations in oral squamous cell carcinoma from Japanese patients by targeted amplicon sequencing. *Oncotarget* **8**, 59113 (2017).
47. Ahmad, H. et al. Mutational landscape and in-silico analysis of TP53, PIK3CA, and PTEN in patients with breast cancer from khyber Pakhtunkhwa. *ACS Omega* **8**, 43318–43331 (2023).
48. Goode, E. L., Ulrich, C. M. & Potter, J. D. J. Polymorphisms in DNA repair genes and associations with cancer risk. *Cancer Epidemiol. Biomarkers Prev.* **11**, 1513–1530 (2002).
49. Negrini, S., Gorgoulis, V. G. & Halazonetis, T. D. J. Genomic instability—an evolving hallmark of cancer. *Nat. Rev. Mol. Cell Biol.* **11**, 220–228 (2010).
50. Xu, Y., Wang, H., Nussinov, R. & Ma, B. J. P. Protein charge and mass contribute to the spatio-temporal dynamics of protein–protein interactions in a minimal proteome. *Proteomics* **13**, 1339–1351 (2013).
51. Sim, N.-L. et al. SIFT web server: Predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res.* **40**, W452–W457 (2012).
52. Adzhubei, I., Jordan, D. M. & Sunyaev, S. R. J. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet.* **76**, 7–20 (2013).
53. Seelow, D. & Robinson, P. Variant pathogenicity prediction. *Comput. Exome Genome Anal.*, 329–345 (Chapman and Hall/CRC, 2017).
54. Castellana, S., Fusilli, C. & Mazza, T. J. A broad overview of computational methods for predicting the pathophysiological effects of non-synonymous variants. *Data Min. Tech. Life Sci.,* 423–440 (2016).
55. Hassan, M. S., Shaalan, A., Dessouky, M., Abdelnaiem, A. E. & ElHefnawi, M. J. G. Evaluation of computational techniques for predicting non-synonymous single nucleotide variants pathogenicity. *Genomics* **111**, 869–882 (2019).
56. Zemla, A. et al. Genesv–An approach to help characterize possible variations in genomic and protein sequences. *Bioinform. Biol. Insights* **8**, BBI.S13076 (2014).
57. Maitra, A. et al. Mutational landscape of Gingivo-Buccal oral squamous cell carcinoma reveals new recurrently-mutated genes and molecular subgroups. *Nat. Commun.* **4**, 2873 (2013).
58. Hyodo, T. et al. The mutational spectrum in whole exon of p53 in oral squamous cell carcinoma and its clinical implications. *Sci. Rep.* **12**, 21695 (2022).
59. Jing, C. et al. Next-generation sequencing-based detection of EGFR, KRAS, BRAF, NRAS, PIK3CA, Her-2 and TP53 mutations in patients with non-small cell lung cancer. *Mol. Med. Rep.* **18**, 2191–2197 (2018).
60. Cho, Y., Gorina, S., Jeffrey, P. D. & Pavletich, N. P. J. S. Crystal structure of a p53 tumor suppressor-DNA complex: Understanding tumorigenic mutations. *Science* **265**, 346–355 (1994).
61. Kitayner, M. et al. Structural basis of DNA recognition by p53 tetramers. *Mol. Cell* **22**, 741–753 (2006).
62. Poeta, M. L. et al. TP53 mutations and survival in squamous-cell carcinoma of the head and neck. *N. Engl. J. Med.* **357**, 2552–2561 (2007).
63. Kastan, M. B. & Bartek, J. J. N. Cell-cycle checkpoints and cancer. *Nature* **432**, 316–323 (2004).
64. Ciccia, A. & Elledge, S. J. The DNA damage response: Making it safe to play with knives. *Mol. Cell* **40**(2), 179–204 (2010).
65. Jette, N. R. et al. ATM-deficient cancers provide new opportunities for precision oncology. *Cancers* **12**, 687 (2020).
66. Villaruz, L. C. et al. ATM protein is deficient in over 40% of lung adenocarcinomas. *Oncotarget* **7**, 57714 (2016).
67. Ricciuti, B. et al. Clinicopathologic, genomic, and immunophenotypic landscape of ATM mutations in non–small cell lung cancer. *Clin. Cancer Res.* **29**, 2540–2550 (2023).
68. Choi, M., Kipps, T. & Kurzrock, R. J. ATM mutations in cancer: Therapeutic implications. *Mol. Cancer Ther.* **15**, 1781–1791 (2016).

69. Randon, G. et al. Prognostic impact of ATM mutations in patients with metastatic colorectal cancer. *Sci. Rep.* **9**, 2858 (2019).
70. Rundle, S., Bradbury, A., Drew, Y. & Curtin, N. J. J. C. Targeting the ATR-CHK1 axis in cancer therapy. *Cancers* **9**, 41 (2017).
71. Fang, Y. et al. ATR functions as a gene dosage-dependent tumor suppressor on a mismatch repair-deficient background. *EMBO J.* **23**, 3164–3174 (2004).
72. Brown, E. J. & Baltimore, D. J. ATR disruption leads to chromosomal fragmentation and early embryonic lethality. *Genes Dev.* **14**, 397–402 (2000).
73. Bertoni, F. et al. CHK1 frameshift mutations in genetically unstable colorectal and endometrial cancers. *Genes Chromosomes Cancer* **26**, 176–180 (1999).
74. Vassileva, V., Millar, A., Briollais, L., Chapman, W. & Bapat, B. J. Genes involved in DNA repair are mutational targets in endometrial cancers with microsatellite instability. *Cancer Res.* **62**, 4095–4099 (2002).
75. Menoyo, A. et al. Somatic mutations in the DNA damage-response genes ATR and CHK1 in sporadic stomach tumors with microsatellite instability. *Cancer Res.* **61**, 7727–7730 (2001).
76. Consortium, A. P. G. *et al.* AACR Project GENIE: Powering precision medicine through an international consortium. *Cancer Discov.* **7**, 818-831 (2017).
77. Bartek, J. & Lukas, J. J. Chk1 and Chk2 kinases in checkpoint control and cancer. *Cancer Cell* **3**, 421–429 (2003).
78. Bell, D. W. et al. Heterozygous germ line hCHK2 mutations in Li-Fraumeni syndrome. *Science* **286**, 2528–2531 (1999).
79. Cybulski, C. et al. CHEK2 is a multiorgan cancer susceptibility gene. *Hum. Genet.* **75**, 1131–1135 (2004).

Acknowledgements

We acknowledge the financial support obtained through Higher Education Commission Pakistan. We are also thankful to Dr. Muslim Khan from KCD for helping in the collection of the samples This research was conducted as part of the project titled "ICRG-46. Diagnostic, prognostic and predictive biomarkers for oral squamous cell carcinoma" in collaboration with the University of Sheffield, UK. The Pakistani part of the project was funded by the Higher Education Commission (HEC) of Pakistan, while the UK part was supported by the British Council

Author contributions

WN, ATK, AA, SAK, MTS conceptualized the project; WN, FN, HA performed the lab procedures and recruitment; MK assisted in sample collection; IAK and MI facilitated in sequencing; ATK, HA, WN, MF performed the bioinformatics; WN, ATK, FN drafted the manuscript; MA, AA, SAK, DAG reviewed and improved the manuscript, AA and SAK acquired the funding,

Funding

This research was conducted as part of the project titled "ICRG-46. Diagnostic, prognostic and predictive biomarkers for oral squamous cell carcinoma" in collaboration with the University of Sheffield, UK. The Pakistani part of the project was funded by the Higher Education Commission (HEC) of Pakistan, while the UK part was supported by the British Council.

Declarations

Competing interest

The authors declare no competing interests.

Ethical approval

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. The research performed is in accordance with the Declaration of Helsinki.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-91700-x.

Correspondence and requests for materials should be addressed to A.T.K., S.A.K. or A.A.