Cell. Mol. Biol. (ISSN: 1165-158X)

Journal Homepage: www.cellmolbiol.org

Cellular and Molecular Biology

Original Article

Identification of novel variants with predicted pathogenicity as key targets in esophageal cancer

Waqas Ahmad Abbasi[1], Sajida Qureshi[1,*], Muhammad Asif Qureshi[2], Mohammad Saeed Quraishy[1]

[1] *Dow Medical College, Dow University of Health Sciences, Karachi, Pakistan*
[2] *Dow International Medical College, Dow University of Health Sciences, Karachi, Pakistan*
[2] *Department of Pathology, Dow International Medical College, Dow University of Health Sciences Karachi, Karachi, Pakistan*

Article Info

Article history:

Received: May 04, 2025
Accepted: August 22, 2025
Published: September 30, 2025

Use your device to scan and read the article online

Abstract

Esophageal cancer (EC) remains a major global health challenge due to its aggressive nature and poor prognosis. Genetic alterations play a crucial role in tumor progression; however, a deeper understanding of the genetic landscape of EC is essential for identifying novel and potent therapeutic targets. This study aims to identify key genes and their variants with potential pathogenicity driving EC progression. Whole-exome sequencing (WES) was performed on EC samples to identify missense variants. A comprehensive in-silico analysis was conducted using SIFT, FATHMM, PROVEAN, MutationTaster, and LRT to classify high-risk variants. Gene expression, mutation frequency, and prognostic relevance were analyzed using GEPIA and cBioPortal platforms. Protein stability was assessed with MuPro and I-Mutant to evaluate the impact of the identified variant, while protein-protein interaction (PPI) analysis via STRING and enrichment analysis through Metascape were performed to explore associated biological pathways. A total of 331 novel high-risk missense variants were identified across 274 genes and systematically refined, narrowing down to 23 prognostically significant variants in 11 genes (PSMC1, SCN8A, HNRNPA3, RPL23, COL5A2, TBL1XR1, TCP1, HNRNPD, CALM2, ABCC2, and HNRNPA1), which were also among the most differentially expressed in EC. Variants in these genes were predicted to destabilize their corresponding proteins, contributing to EC progression. In-silico survival analysis further indicated significantly worse outcomes for patients harboring alterations in these genes, including others. Protein stability analysis confirmed their destabilizing effects, while functional enrichment highlighted their involvement in key pathways driving tumorigenesis. This study identified 11 key DEGs harboring potentially pathogenic novel missense variants, highlighting vulnerabilities for precision-targeted therapies in EC.

Keywords: Esophageal cancer, Single-nucleotide variants, Potentially pathogenic variants, Therapeutic targets.

1. Introduction

Esophageal cancer (EC), the eleventh most common cancer worldwide in absolute numbers, disproportionately affects middle and low-income countries, with Asia bearing the highest incidence and mortality rates [1-3]. Projections indicate a 63% increase in cases and a 72% rise in mortality by 2040, highlighting the urgent need for improved diagnostic, prognostic, and therapeutic strategies [4]. Despite advancements in multimodal treatment, including chemotherapy, chemoradiotherapy, surgical esophagogastrectomy, and adjuvant immunotherapy, EC remains associated with poor survival outcomes, largely due to limited progress in molecularly targeted treatment options. A deeper understanding of the genetic landscape of EC is essential for identifying novel and potent therapeutic targets, particularly those influenced by genetic variations driving tumorigenesis.

Genetic variations, especially single nucleotide variants (SNVs), play a critical role in EC susceptibility. These alterations contribute significantly to genomic instability and disease progression [5]. While environmental factors are key drivers of EC, studies also suggest a hereditary component, with familial aggregation and segregation analyses indicating a genetic predisposition**.** Among these, missense variants, a type of nsSNV, result in amino acid substitutions, which can alter protein structure, stability, or interactions, thereby influencing functional diversity and disease risk**.** These alterations arise from specific base pair changes at the genomic level, demonstrating their broader role in genomic instability and tumorigenesis in EC.

Among the two histological subtypes, esophageal squamous cell carcinoma (ESCC) is the predominant form globally and in South Asia [6]. Yet, most cases are presented at an advanced, unresectable stage with poor prognosis due to the lack of effective early detection and prognostic biomarkers [7, 8]. Similarly, the molecular mechanisms

* Corresponding author.
E-mail address: sajida.qureshi@duhs.edu.pk (S. Qureshi).
Doi: http://dx.doi.org/10.14715/cmb/2025.71.9.11

driving Barrett's esophagus (BE) progression to esophageal adenocarcinoma (EAC) remain poorly defined, with no clinically validated markers for early diagnosis or disease stratification [9]. This study aims to characterize key genes and their variants driving EC progression, through *in-vitro* and *in-silico* analyses, identifying potential therapeutic targets.

2. Materials and Methods

2.1. Study setting

The study was conducted at Dow University of Health Sciences (DUHS), Karachi, Pakistan. Patients were recruited at the Department of Upper Gastrointestinal (GI) Surgery, Surgery Unit-I, Dr. Ruth K. M. Pfau, Civil Hospital in Karachi, and initial bench work was conducted at the Dow Diagnostic Research and Reference Laboratory. All participants provided informed consent, and the study was approved by the University's Ethical Review Board (DUHS/Approval/2022/818; dated 12th April 2022)

2.2. Endoscopy, biopsy, and data collection

Clinicopathological characteristics of all enrolled patients (n=10) were recorded. Endoscopic tissue samples were obtained using OLYMPUS 160-190 series endoscopes, ensuring precise targeting and minimal tissue damage. Biopsies were conducted under standard clinical protocols, with samples formalin-fixed for histopathology and snap-frozen for nucleic acid extraction. Samples were transported on dry ice and stored at -80°C for long-term preservation. Standard treatment protocols were followed, without any deviations, and patients were followed up for two years from the date of biopsy confirmation to assess survival outcomes.

2.3. DNA extraction

DNA was extracted from frozen tumor samples using the QIAGEN DNA extraction kit, following the manufacturer's protocols. The extracted DNA was dissolved in 50-100 µl of elution buffer, and the DNA concentrations were measured using a NanoDrop spectrophotometer.

2.4. Whole exome sequencing (WES)

Whole-exome sequencing (WES) and initial data analysis were outsourced to Macrogen as a commercial service. Extracted genomic DNA was used to generate exome capture libraries using SureSelect V6 kit. Sequencing by synthesis was performed on the Illumina platform NovaSeq 6000. Paired-end sequences that were generated from the Illumina sequencing platform were mapped to the human reference genome (hg38) using the mapping program Burrows-Wheeler Aligner. PCR duplicates were removed using MarkDUplicates.jar from the "Picard-tools" package. The Binary Alignment and Map (BAM) files were subsequently recalibrated with Base Quality Score Recalibration (BQSR). SNVs and short indels were identified using the Haplo-typeCaller tool in the Genome Analysis Toolkit (GATK). Detected variants were filtered using GATK's Variant Filtration tool. Annotated variants were generated with SnpEff and further filtered by dbSNP and the 1000 Genomes Project. Additionally, an In-house Macrogen program was used to annotate the identified variants with additional databases, including Exome Sequencing Project 6500 (ESP6500), Clinical Variant Database (ClinVar), Database for Functional Predictions of nsSNPs (dbNSFP), and American College of Medical Genetics and Genomics (ACFMG). Sequencing data have been uploaded to the NCBI Sequence Read Archive (SRA) database (Accession ID: PRJNA1182118)

2.5. Identification criteria for functionally impactful nsSNVs

The potential functional impact of identified nsSNVs was evaluated using five *in-silico* tools incorporated within the annotation service provided by Macrogen: Sorting Intolerant From Tolerant (SIFT), Protein Variation Effect Analyzer (PROVEAN), Likelihood Ratio Test (LRT), MutationTaster, and Functional Analysis through Hidden Markov Models (FATHMM). SIFT predicted whether an amino acid substitution affects protein function by assessing evolutionary conservation, with a cutoff of < 0.5 indicating a damaging effect [10]. Since highly conserved regions are often critical for protein function, PROVEAN evaluated the potential deleterious effect of variants by comparing sequence alignment scores, with a threshold of < -2.5 for deleterious predictions [11]**.** LRT, a statistical method comparing neutral and deleterious mutation models, was used to identify harmful mutations in conserved functional regions, with a p-value of < 0.5 indicating a damaging effect [12]. MutationTaster predicted the disease-causing potential of mutations using a Bayesian classifier, with scores > 0.5 (closer to 1) indicating pathogenic variants [13]**.** FATHMM assessed the functional effects of missense variants by combining sequence conservation with domain-specific pathogenicity weights, and variants with a score < -1.5 were considered deleterious [14].

2.6. Variant nomenclature and prioritization

Genetic variants were prioritized based on their frequency across the cohort (n=10), including only those with ≥2 distinct or recurrent mutations, categorized gene-wise, grouping multiple mutations within their respective genes across the cohort. Annotation adhered to HGVS nomenclature, documenting both nucleotide (HGVS.c) and protein (HGVS.p) changes. To assess clinical relevance, patients harboring these variants within genes were highlighted, incorporating survival outcomes in fraction form by calculating the number of patients harboring mutations in a gene within the cohort (x) to those who did not survive (y).

2.7. In-silico

2.7.1. Genomic alteration analysis

In-silico genomic alteration analysis was conducted using cBioPortal (https://www.cbioportal.org/; accessed February 24, 2025), leveraging publicly available datasets to contextualize our findings at the genomic/variant levels. Data from ESCC and EAC were analyzed, sourced from the International Cancer Genome Consortium (ICGC), Nature 2014 (n=88), and The Cancer Genome Atlas (TCGA), Firehose Legacy (n=186), respectively, totaling n=274 patient samples.

The OncoPrint tool was used to visualize genomic alterations, displaying data as event per patient (alteration data available, n=273). Survival analysis was performed using the Kaplan-Meier (KM) estimator, comparing altered vs. unaltered groups across queried genes. A p-value < 0.05 was considered statistically significant.

2.7.2. Expression analysis

Gene expression analysis was conducted using the Gene Expression Profiling Interactive Analysis (GEPIA) database (http://gepia.cancer-pku.cn/; accessed February 25, 2025), a publicly available repository containing RNA expression data from 9,736 tumors and 8,587 normal samples across all TCGA cancers. The 'Multiple Gene Comparison' feature was utilized to evaluate expression patterns, with tumor samples from EC compared against their matched normal counterparts. To further assess differential expression, log2 fold change (Log2FC) values were calculated, where log2FC > 1 indicated oncogenic potential.

2.7.3. Protein's stability assessment

Protein stability changes were predicted using MUpro and I-Mutant v2.0, both sequence-based machine learning tools. Protein sequences were retrieved from UniProt (https://www.uniprot.org/; accessed on 26 February 2025). MUpro (http://mupro.proteomics.ics.uci.edu/; accessed on 26 February 2025) utilizes Support Vector Machines (SVM) and Neural Networks (NN) to assess stability alterations without requiring tertiary structure information. I-Mutant v2.0 (https://folding.biofold.org/i-mutant/i-mutant2.0.html; accessed on 26 February 2025) is an SVM-based predictor operating at pH 7 and 25°C. The Free Energy change value (ΔΔG) function was used to evaluate the effect of mutations, where $\Delta\Delta G < 0$ indicates a decrease in protein stability, and $\Delta\Delta G > 0$ indicates an increase.

2.7.4. Protein network and functional enrichment analysis

Protein-protein interaction (PPI) networks were constructed using STRING (http://string-db.org; accessed on 26 Feb 2025) with a minimum interaction score of 0.4 (medium confidence). Edges were based on multiple evidence sources, including text mining, experimental data, curated databases, co-expression, neighborhood, gene fusion, and co-occurrence. Functional enrichment analysis was performed using Metascape (https://metascape.org; accessed on 26 Feb 2025) in "Custom Analysis" mode. Gene Ontology (GO) terms for Biological Process (BP), Cellular Component (CC), and Molecular Function (MF) were analyzed alongside pathway databases, including the Kyoto Encyclopedia of Genes and Genomes (KEGG), Reactome, Wiki-Pathways, Protein ANalysis THrough Evolutionary Relationships (PANTHER), and the comprehensive resource of mammalian protein complexes (CORUM). A significance threshold of $p < 0.01$ and a minimum enrichment factor of 1.5 was applied.

2.8 Statistical analysis

Data was analyzed using SPSS version 26. Quantitative data such as age was expressed as mean with standard deviation. Categorical variables, including gender, histopathology, tumor site, grade of differentiation, clinical TNM staging, tumor length and surgery, were reported as frequencies and percentages.

3. Results

3.1. Clinicopathological characteristics

All patients (n=10) were presented with clinical symptoms of esophageal malignancy at the Department of Upper GI, Surgical Unit-1, Dr. Ruth K. M. Pfau Civil Hospital Karachi. A diagnostic workup was conducted following standard clinical protocols, and the mean age of the cohort was 51.7 years. Of the patients, 60% (n=6) were male and 40% (n=4) were female. Histopathological analysis confirmed EC, with ESCC being the predominant type (90%, n=9), with a single case of EAC (10%, n=1). Tumors were primarily located in the lower thoracic esophagus (60%, n=6), followed by the mid-thoracic (20%, n=2) and upper thoracic regions (20%, n=2).

Tumor differentiation showed that 50% (n=5) of the cases were moderately differentiated, 30% (n=3) were well-differentiated, and 20% (n=2) were poorly differentiated. Clinical staging revealed 50% (n=5) had cT3, 40% (n=4) had cT4, and 10% (n=1) had cT2. Nodal involvement was observed in 20% (n=2) of cases, marked by regional nodal metastasis (N1), while 80% (n=8) had no nodal involvement (N0). In terms of metastatic spread, 90% (n=9) had no distant metastasis (M0), while only 10% (n=1) showed evidence of distant metastasis (M1).

Most patients (60%, n=6) were classified as stage II, 10% (n=1) as stage III, and 30% (n=3) as stage IV. Tumor sizes were also categorized as follows: <5 cm in 30% (n=3), 5–10 cm in 60% (n=6), and >10 cm in 10% (n=1) of the cases. Additionally, surgery was performed in 20% (n=2) of patients (lap McKeon Esophagectomy), while 80% (n=8) did not undergo curative surgery, primarily due to tumor unresectability and/or complications noted during follow-up assessments.

3.2. Classification of potentially pathogenic and novel missense variants

WES was performed on all samples (n=10) with high sequencing quality, achieving Q20 and Q30 scores of >97% and >92%, at 100× coverage, ensuring robust variant detection. Each sample was assigned a unique patient ID prior to analysis.

A total of 912,214 SNVs were identified across the cohort (Table 1). The number of SNVs identified per patient was relatively consistent, with an average of 91,221 SNVs (range: 82,637–102,165), indicating no extreme variation among samples. This suggests a uniform sequencing depth and variant detection across patients.

To determine the potential impact of these variants, we focused on missense variants, which can alter protein structure and function, thereby contributing to disease progression. We applied multiple *in-silico* prediction algorithms, including SIFT, LRT, MutationTaster, FATHMM, and PROVEAN, to evaluate the potential pathogenicity of variants. Using established pathogenicity thresholds (SIFT <0.5, PROVEAN <-2.5, LRT <0.5, MutationTaster >0.5, FATHMM <-1.5), 703 missense variants were classified as potentially deleterious/damaging, hence contributing towards pathogenicity.

To assess novelty, these variants were cross-referenced with the Single Nucleotide Polymorphism database (dbSNP). This filtering shortlisted 331 novel missense variants that have not been previously reported, suggesting them as potential candidate variants/targets for further investigation in EC.

3.3. Gene-wise categorization of novel potentially pathogenic variants and their prognostic implications

All 331 novel potentially pathogenic variants were classified according to their respective genes to facilitate a

Table 1. Summary of identified variants, potentially pathogenic missense variants, and novel variants among them across the cohort

Pt. IDs	Total identified Variants	Potential Pathogenic Missense Variants (SIFT, LRT, Mutation Taster, FATHMM, PROVEAN)	Novel Potentially Pathogenic Missense Variants
EC1	96453	60	16
EC2	94506	79	36
EC3	102165	93	51
EC4	85503	51	17
EC5	89010	52	37
EC6	84087	82	44
EC7	82637	51	21
EC8	91623	64	27
EC9	88846	74	33
EC10	97384	97	49
Total	912214	703	331

more meaningful interpretation of the data. Variants detected across the cohort were grouped gene-wise to provide a comprehensive overview of the genes harboring these novel alterations (Supplementary File 1). Only those genes containing ≥2 distinct or recurrent variants across the cohort (n=10) were prioritized to refine the analysis. This categorization (Table 2) incorporated both nucleotide-level and protein-level alterations, resulting in the identification of 40 genes harboring ≥2 variants each.

To further explore the relevance of these variants at the gene level, two-year survival status (alive/deceased) was mapped for all patients bearing variants within each gene. While some patients had a single variant per gene, others had multiple variants. In our cohort, 7 out of 10 patients (70%) did not survive in a 2-year follow-up from the biopsy date. The survival fraction for each gene in Table 2 was determined by calculating the number of patients harboring mutations in that gene within the cohort (x), to the who did not survive (y). Only genes in which all affected patients (x/y = 100% death) did not survive were considered to be potentially interfering with prognosis. Through this approach, 23 key genes were identified *(PSMC1, HOMER2, HOXA13, SCN8A, LRP2, HNRNPA3, PSMC5, MTMR2, RPL23, ITPR1, COL5A2, TBL1XR1, ITGB1, ZFAND1, PBX2, TCP1, HNRNPD, CALM2, PCBD2, CACNA1E, ABCC2, AFG3L2, and HNRNPA1)* as important prognostic indicators, highlighting their significance in EC.

3.4. In-silico prognostic impact of key genetic alterations via cBioPortal

Building upon our cohort-based findings and to further investigate the genomic landscape of EC in detail, we analyzed publicly available datasets to assess alterations in our 23 key genes across an EC patient cohort (n=273). Genomic alterations included amplifications, deep deletions, splice site mutations, truncating mutations, and missense mutations, all classified as variants of unknown significance.

Our analysis revealed that all 23 genes exhibited alterations within the EC cohort, with some genes demonstrating a higher frequency of alterations. *TBL1XR1* was the most frequently altered gene, observed in 17% of patients (n=273), followed by *CACNA1E* in 9%, and *ITPR1* and *HOXA13* in 5% of cases. *SCN8A*, *RPL23*, *ZFAND1*, and *PBX2* were each altered in 4% of patients, while the remaining genes were altered in <3% of cases (Figure 1A).

To assess the prognostic relevance of these alterations, we performed a KM survival analysis for the 23 key genes as a collective cluster within the cohort (n=273). Among these cases, 129 patients had no detected alterations, whereas 144 patients harbored at least one alteration in the selected key genes. Of the altered group, 65 patients met the event (expired). Survival analysis comparing the altered and unaltered groups demonstrated that patients without alterations in these 23 genes had significantly better overall survival than those with alterations. The difference in survival was statistically significant (log-rank p = 0.01), indicating that alterations in these genes are associated with a worse prognosis in EC (Figure 1B).

3.5. Differential expression of key genes in EC

To further assess the expression patterns of the 23 key genes, we analyzed publicly available transcriptomic data using log2 (TPM + 1) transformation to compare expres-

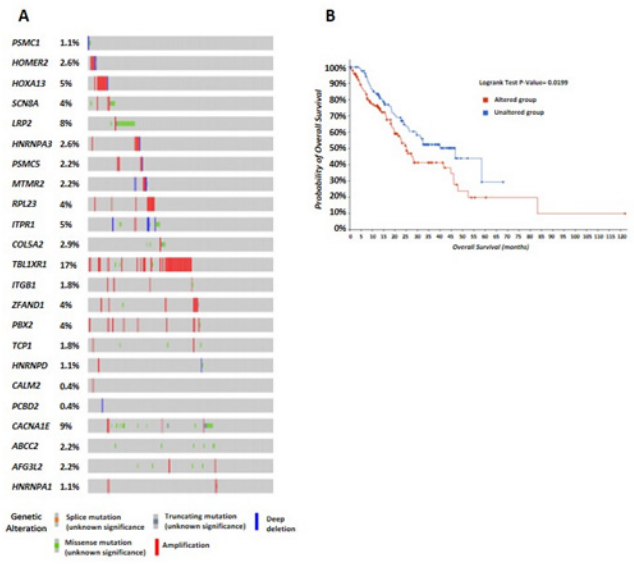

Fig. 1. Genomic alterations and prognostic impact of 23 key genes in EC via cBioPortal. (A) Oncoprint showing the frequency and types of genetic alterations in the 23 key genes. **(B)** Kaplan-Meier survival analysis comparing overall survival between the altered group (red) and unaltered group (blue), showing a trend of worse survival in patients with gene alterations in the identified key genes

Table 2. Genes harboring ≥ 2 distinct or recurrent novel potentially pathogenic missense variants across the cohort, ranked by variant frequency, with gene-wise survival data.

Gene	cDNA Change (HGVS.c)	Amino Acid Substitution (HGVS.p)	Survival: No. of pt. bearing alteration in gene (x) / died (y)
RPS4X	c.668C>T: 1, c.665T>C: 1, c.476C>T: 1, c.452A>G: 1, c.433C>T: 1, c.454C>T: 1, c.442C>T: 1	p.Ser223Phe: 1, p.Leu222Pro: 1, p.Thr159Ile: 1, p.Asp151Gly: 1, p.Arg145Cys: 1, p.Pro152Ser: 1, p.Arg148Cys: 1	3/1.
PSMC1	c.1141G>C: 1, c.980T>C: 1, c.791C>T: 1, c.803G>T: 1	p.Asp381His: 1, p.Val327Ala: 1, p.Pro264Leu: 1, p.Arg268Leu: 1	3/3; 100% DEATH
HOMER2	c.221T>C: 1, c.218A>C: 1, c.188C>A: 1, c.176G>A: 1	p.Phe74Ser: 1, p.Lys73Thr: 1, p.Pro63Gln: 1, p.Ser59Asn: 1	1/1; 100% DEATH
PPAT	c.1049A>C: 2, c.1247G>C: 1, c.1238T>A: 1, c.1051C>T: 1	p.Asn350Thr: 2, p.Arg416Pro: 1, p.Val413Glu: 1, p.Arg351Trp: 1	3/2.
TPM3	c.317G>A: 1, c.523G>A: 1, c.722A>G: 1	p.Arg106His: 1, p.Gly175Arg: 1, p.Glu241Gly: 1	3/2.
RANBP9	c.1358T>G: 1, c.1355T>G: 1, c.1352A>G: 1	p.Val453Gly: 1, p.Met452Arg: 1, p.Glu451Gly: 1	1/0.
NALCN	c.2846T>C: 1, c.3128T>A: 1, c.3125A>T: 1	p.Val949Ala: 1, p.Leu1043Gln: 1, p.Lys1042Ile: 1	2/1.
PTPN11	c.1493G>A: 1, c.194T>C: 1, c.197A>T: 1	p.Arg498Gln: 1, p.Leu65Pro: 1, p.Tyr66Phe: 1	2/1.
ARIH2	c.1280C>T: 1, c.1282T>A: 1, c.1291G>A: 1	p.Thr427Ile: 1, p.Tyr428Asn: 1, p.Ala431Thr: 1	1/0.
BMPR1B	c.1222A>C: 1, c.1571T>A: 1	p.Met408Leu: 1, p.Met524Lys: 1	1/0.
HOXA13	c.1130A>G: 1, c.1129G>A: 1	p.Glu377Gly: 1, p.Glu377Lys: 1	1/1; 100% DEATH
PRKAA2	c.1213G>C: 1, c.1214G>T: 1	p.Gly405Arg: 1, p.Gly405Val: 1	1/0.
SCN8A	c.4496A>G: 1, c.4502A>G: 1	p.Lys1499Arg: 1, p.Gln1501Arg: 1	1/1; 100% DEATH
HDAC1	c.785G>A: 1, c.1016C>A: 1	p.Gly262Asp: 1, p.Pro339Gln: 1	2/1.
LRP2	c.4058G>T: 1, c.1575T>G: 1	p.Cys1353Phe: 1, p.Phe525Leu: 1	2/2; 100% DEATH
HNRNPA3	c.502G>A: 1, c.523G>A: 1	p.Gly168Arg: 1, p.Asp175Asn: 1	1/1; 100% DEATH
FECH	c.856G>T: 1, c.847C>T: 1	p.Val286Leu: 1, p.Pro283Ser: 1	1/0.
PSMC5	c.215A>G: 1, c.220G>A: 1	p.Tyr72Cys: 1, p.Gly74Arg: 1	1/1; 100% DEATH
MTMR2	c.554C>T: 1, c.548C>G: 1	p.Pro185Leu: 1, p.Ala183Gly: 1	1/1; 100% DEATH
ABCB1	c.1735G>A: 1, c.1730G>C: 1	p.Gly579Ser: 1, p.Arg577Thr: 1	1/0.
RPL23	c.23G>A: 1, c.16C>T: 1	p.Gly8Glu: 1, p.Arg6Cys: 1	1/1; 100% DEATH
ITPR1	c.7232T>A: 1, c.7241T>G: 1	p.Phe2411Tyr: 1, p.Val2414Gly: 1	1/1; 100% DEATH
COL5A2	c.1133G>T: 1, c.3001C>G: 1	p.Gly378Val: 1, p.Arg1001Gly: 1	2/2; 100% DEATH
TBL1XR1	c.1451G>A: 1, c.1439A>G: 1	p.Gly484Asp: 1, p.Tyr480Cys: 1	1/1; 100% DEATH
ITGB1	c.128C>A: 1, c.127C>A: 1	p.Pro43Gln: 1, p.Pro43Thr: 1	1/1; 100% DEATH
ZFAND1	c.235T>C: 1, c.205T>C: 1	p.Cys79Arg: 1, p.Cys69Arg: 1	1/1; 100% DEATH
PBX2	c.896G>A: 1, c.880T>C: 1	p.Arg299Lys: 1, p.Trp294Arg: 1	1/1; 100% DEATH
TCP1	c.1489G>C: 1, c.1034T>C: 1	p.Asp497His: 1, p.Leu345Ser: 1	2/2; 100% DEATH
HNRNPD	c.686C>T: 1, c.675C>G: 1	p.Thr229Ile: 1, p.Phe225Leu: 1	1/1; 100% DEATH
CALM2	c.470T>C: 1, c.464G>A: 1	p.Val157Ala: 1, p.Arg155His: 1	1/1; 100% DEATH
PEX5	c.428T>G: 1, c.429T>A: 1	p.Phe143Cys: 1, p.Phe143Leu: 1	1/0.
PCBD2	c.322G>A: 1, c.338C>T: 1	p.Asp108Asn: 1, p.Thr113Ile: 1	1/1; 100% DEATH
CACNA1E	c.4182C>G: 1, c.3626G>A: 1	p.Ser1394Arg: 1, p.Gly1209Asp: 1	2/2; 100% DEATH
FOXO1	c.1956G>C: 1, c.1943C>G: 1	p.Trp652Cys: 1, p.Thr648Arg: 1	1/0.
GPD2	c.443T>C: 1, c.451G>T: 1	p.Leu148Pro: 1, p.Ala151Ser: 1	1/0.
KARS	c.1391C>T: 1, c.1803G>A: 1	p.Pro464Leu: 1, p.Met601Ile: 1	2/1.
ABCC2	c.4186G>T: 1, c.4187A>C: 1	p.Asp1396Tyr: 1, p.Asp1396Ala: 1	1/1; 100% DEATH
AFG3L2	c.1364G>A: 2	p.Arg455Gln: 2	2/2; 100% DEATH
CHCHD2	c.350A>G: 2	p.Glu117Gly: 2	2/0.
HNRNPA1	c.547A>C: 2	p.Lys183Gln: 2	2/2; 100% DEATH

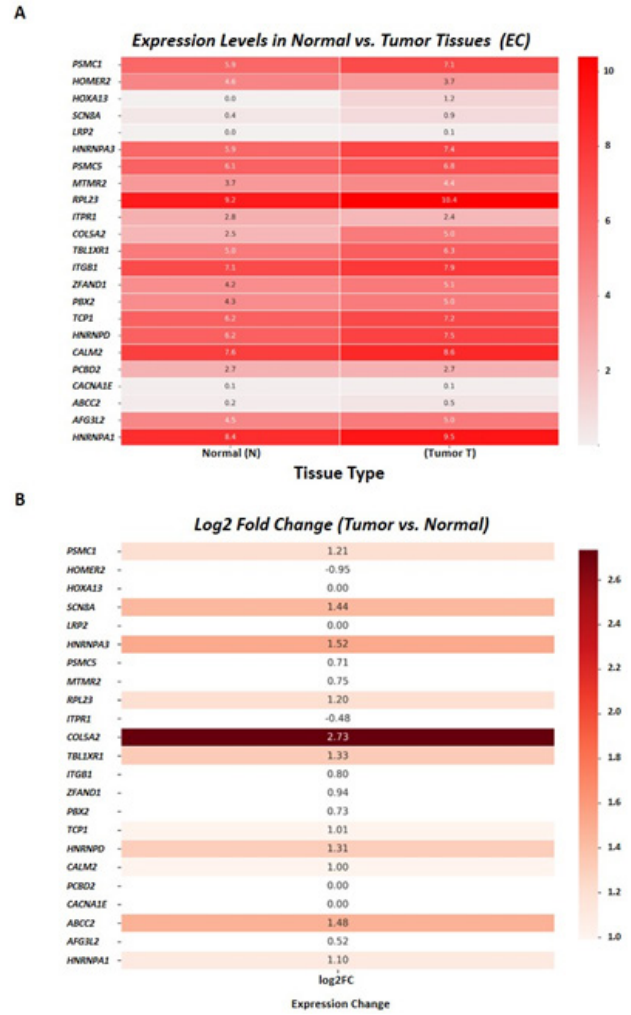

Fig. 2. Differential expression analysis of 23 key genes in EC. (A) mRNA expression levels of key genes in normal (N) and tumor (T) tissues, based on transcriptomic data [log2(TPM +1) transformation]. **(B)** Log2 fold change (log2FC) values comparing tumor and normal tissues, sorting differentially expressed genes (DEGs) with log2FC > 1, indicating potential oncogenic involvement in EC.

sion levels between tumor and matched normal tissues in EC. Differential expression was observed across multiple genes (Figure 2A). To further refine this outcome, we calculated log2FC values, applying a cutoff of log2FC > 1 to identify genes with significant upregulation, indicative of potential oncogenic relevance. This approach highlighted 11 key differentially expressed genes (DEGs) with notable differential expression, among which *COL5A2* exhibited the highest upregulation (log2FC = 1.78), followed by *HNRNPA3* (log2FC = 1.52), *ABCC2* (log2FC = 1.48), *SCN8A* (log2FC = 1.44), *TBL1XR1* (log2FC = 1.33), and *HNRNPD* (log2FC = 1.31). The remaining genes demonstrated log2FC > 1 but ≤ 1.3, reinforcing their fair potential for oncogenic involvement in EC (Figure 2B).

3.6. Variant-induced protein destabilization and pathway involvement

We assessed the impact of each identified variant on protein stability among the refined 11 key DEGs within EC tissues, focusing on changes in Gibbs free energy (ΔΔG), which influence protein folding and stability. Our analysis showed that several variants consistently led to destabilization across all predictive models and hence were potent. Specifically, variants in *PSMC1* (V327A, P264L), *SCN8A* (Q1501R), *HNRPA3* (D175N), *RPL23* (R6C), *TCP1* (L345S, D497H), *HNRNPD* (T229I, F225L), *CALM2* (V157A, R155H), *ABCC2* (D1396Y, D1396A), and *HNRNPA1* (K183Q) exhibited negative ΔΔG values across all applied algorithms, indicating compromised structural stability (Table 3). While other variants exhibited varying degrees of stability loss across different models, most remained consistently on the destabilized negative end, suggesting a potential impact on protein function.

To understand how these structurally compromised proteins influence cellular dynamics, we mapped their interactions through a PPI network using STRING. This revealed significant connections among the 11 key DEGs, forming a network of 11 nodes and 7 edges, with an average node degree of 1.27, an average local clustering coefficient of 0.242, and a PPI enrichment p-value of 0.04 (Figure 3A). These findings highlight a non-random association between these proteins, emphasizing their potential functional interplay in EC.

Further enrichment analysis via Metascape for these DEGs provided deeper insights into their critical role, highlighting enrichment in key pathways such as positive regulation of telomere maintenance via telomerase (GO:0032212), axon guidance (R-HSA-422475), metabolism of RNA (R-HSA-8953854), organelle biogenesis and maintenance (R-HSA-1852241), cellular responses to stimuli (R-HSA-8953897), and transport of small molecules (R-HSA-382551) (Figure 3B).

4. Discussion

This study identified 331 novel missense variants with potential pathogenicity (deleterious/damaging) across 274 genes in our cohort (S1). Among these, 23 variants across 11 DEGs exhibited prognostic relevance, with their

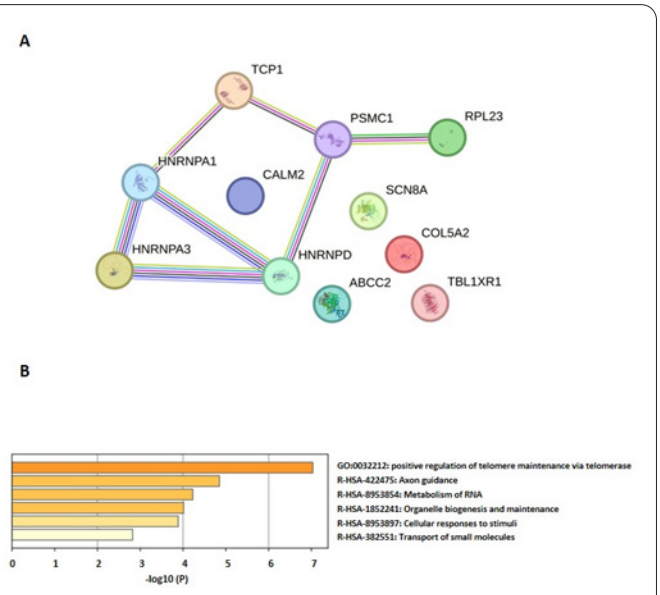

Fig. 3. Protein-protein interaction (PPI) network and functional enrichment analysis of 11 key differentially expressed genes (DEGs) in EC. (A) PPI network of the 11 key DEGs, constructed using STRING. Nodes represent proteins, while edges indicate predicted interactions, forming a network of 11 nodes and 7 edges. **(B)** Functional enrichment analysis of the 11 DEGs using Metascape. The bar plot illustrates significantly enriched pathways, with statistical significance represented as -log10(P) values on the y-axis. Higher log10(P) values indicate lower p-values, reflecting stronger pathway enrichment.

Table 3. Predicted impact of free energy changes on protein stability for key missense variants in 11 EC-related DEGs identified in our cohort.

	Gene	Amino Acid Change	MUpro ΔΔG (kcal/mol)	MUpro_SVM[a] ΔΔG (kcal/mol)	MUpro_NN[b] ΔΔG (kcal/mol)	I-Mutant ΔΔG (kcal/mol)
1	***PSMC1*** **(UniProt ID: P62191)**	D381H	-1.38 (DS)	-0.93 (DS)	-0.85 (DS)	0.32 (IS)
		V327A	-1.59 (DS)	-0.82 (DS)	-0.99 (DS)	-2.47 (DS)
		P264L	-0.45 (DS)	-0.70 (DS)	-0.80 (DS)	-1.26 (DS)
		R268L	0.29 (IS)	0.08 (IS)	0.75 (IS)	-0.72 (DS)
2	***SCN8A*** **(UniProt ID: Q9UQD0-1)**	K1499R	-0.26 (DS)	0.04 (IS)	-0.69 (DS)	0.08 (IS)
		Q1501R	-0.50 (DS)	-0.13 (DS)	-0.53 (DS)	-0.06 (DS)
3	***HNRNPA3*** **(UniProt ID: P51991)**	G168R	-0.83 (DS)	0.18 (IS)	0.64 (IS)	-1.88 (DS)
		D175N	-1.31 (DS)	-1 (DS)	-0.99 (DS)	-1.53 (DS)
4	***RPL23*** **(UniProt ID: P51991)**	G8E	-0.41 (DS)	-0.38 (DS)	-0.82 (DS)	0.18 (IS)
		R6C	-1.01 (DS)	-0.59 (DS)	-0.95 (DS)	-0.88 (DS)
5	***COL5A2*** **(UniProt ID: P05997)**	G378V	-1.26 (DS)	-0.29 (DS)	0.61 (IS)	-0.12 (DS)
		R1001G	-0.53 (DS)	0.37 (IS)	-0.79 (DS)	-0.17 (DS)
6	***TBL1XR1*** **(UniProt ID: Q9BZK7)**	G484D	-0.44 (DS)	1 (IS)	0.78 (IS)	-1.49 (DS)
		Y480C	-0.99 (DS)	-0.38 (DS)	-0.90 (DS)	0.75 (IS)
7	***TCP1*** **(UniProt ID: P17987)**	D497H	-1.24 (DS)	-0.99 (DS)	-0.99 (DS)	-1.02 (DS)
		L345S	-1.50 (DS)	-1 (DS)	-0.92 (DS)	-2.01 (DS)
8	***HNRNPD*** **(UniProt ID: Q14103)**	T229I	-0.97 (DS)	-0.85 (DS)	-0.98 (DS)	-1.68 (DS)
		F225L	-0.73 (DS)	-0.38 (DS)	-0.89 (DS)	-1.98 (DS)
9	***CALM2*** **(NCBI ID: NP_001292553.1)**	V157A	-1.57 (DS)	-1 (DS)	-0.99 (DS)	-1.90 (DS)
		R155H	-1.40 (DS)	-1 (DS)	-0.99 (DS)	-1.20 (DS)
10	***ABCC2*** **(UniProt ID:Q92887)**	D1396Y	-0.99 (DS)	-1 (DS)	-0.99 (DS)	-0.22 (DS)
		D1396A	-1.13 (DS)	-1 (DS)	-0.99 (DS)	-0.87 (DS)
11	***HNRNPA1*** **(UniProt ID:P09651)**	K183Q	-0.25 (DS)	-0.54 (DS)	-0.69 (DS)	-0.64 (DS)

DS: Decreased Stability

variants causing structural instability in the corresponding proteins (Table 3). Such deleterious variants are critical in disease progression, as they destabilize the proteins, disrupt the functional domains, and impair essential cellular processes, often resulting in loss of function or aberrant activity linked to malignancies [15-18].

To identify high-risk variants, an integrative approach was employed using SIFT, LRT, FATHMM, PROVEAN, and MutationTaster. Only variants consistently predicted as deleterious across all tools were classified as high-risk (n=331) (Table 1). Our WES analysis, coupled with annotated data (SRA Accession ID: PRJNA1182118), provides a foundational resource for characterizing deleterious/damaging variants. This dataset serves as an important tool for advancing research on variants in EC and guiding future investigations. However, deriving biologically meaningful insights from such an extensive dataset necessitates systematic refinement. To further contextualize our findings, *in-silico* analyses were integrated; those with publicly available datasets. Given that mRNA activity generally correlates with protein functions, GEPIA and cBioPortal databases were utilized to assess mRNA expression profiles, mutation patterns, and their specific prognostic significance in EC.

In line with this, we performed an *in-silico* survival analysis using publicly available EC datasets to validate 23 key genes that were initially classified as potentially prognostic in our cohort (survival relevance; 100% death) (Table 2). KM survival analysis in cBioPortal demonstrated that patients (n=273) harboring alterations in these genes had significantly worse survival outcomes (log-rank p = 0.01), confirming our prognostic standpoint. Similarly, Liu et al. identified 27 metabolic genes with genomic alterations as independent prognostic factors for EC, including five (*GABRA2, ATP13A1, ABCC2, KCNH8, LPIN3*) specifically with somatic mutations [19]. Although the public dataset in our *in-silico* assessment contains variants within these genes that are prognostically relevant, they differ from the novel variants reported in our study (Table 2). Notably, the major confounding variable among patients in terms of prognosis, which could have been surgical curative resection, was ruled out, as all patients (n=2) who underwent surgery were alive at the end of the 2-year follow-up.

Interestingly, further investigation of expression patterns revealed significant upregulation of several genes in EC among the pinpointed ones, suggesting their potential oncogenic relevance. Notably, 11 DEGs (*PSMC1, SCN8A, HNRNPA3, RPL23, COL5A2, TBL1XR1, TCP1, HNRNPD, CALM2, ABCC2*, and *HNRNPA1*) were identified as potential contributors to EC progression. Structural analysis of the mutations in these genes indicated that they lead to protein destabilization, potentially impairing normal biological functions. Among these, the most impactful variants were those exhibiting excessively decreased stability across all applied algorithms; assessing free ener-

gy changes, including *PSMC1* (V327A, P264L), *SCN8A* (Q1501R), *HNRNPA3* (D175N), *RPL23* (R6C), *TCP1* (L345S, D497H), *HNRNPD* (T229I, F225L), *CALM2* (V157A, R155H), *ABCC2* (D1396Y, D1396A), and *HNRNPA1* (K183Q). While our study primarily focused on somatic alterations and their role in EC progression, many of the above-identified variants may also have germline significance, warranting future exploration. Previous studies have highlighted the importance of inherited mutations in EC susceptibility, such as rare germline mutations in *DNAH9, GKAP1, BAG1, NFX1, FUK*, and *DDOST*, which have been linked to EC risk [20]. Additionally, Golyan et al. identified 22 rare germline variants in familial EC cases, with mutations in *CDK11A, ARID1A, JMJD6, MAML3*, *CDKN2AIP*, and *PHLDA1* implicated in chromatin remodeling and cell-cycle regulation [21].

Most of the 11 key DEGs identified as robust markers in our study remain underexplored in EC in terms of mechanistic understanding, with some having been previously implicated in other cancers or key biological processes. For instance, voltage-gated sodium channels (VGSCs), including *SCN8A*, have been recognized as prognostic biomarkers and potential therapeutic targets in other cancers, yet evidence supporting their role in EC is scarce. VGSCs are of particular interest due to their involvement in enhancing carcinoma cell invasiveness and metastasis [22]. The heterogeneous nuclear ribonucleoproteins (hnRNPs), including *HNRNPA3*, play critical roles in RNA processing. *HNRNPA3* has been linked to pre-mRNA splicing and mRNA transcription, and its overexpression has been associated with poor prognosis in EC [23, 24]. Similarly, elevated expression of *HNRNPD* and *HNRNPA1*, which encode the hnRNP D and A1 proteins, has been reported in ESCC [25]. *RPL23*, an oncogene in gastric cancer, appears to play a similar role in EC, promoting proliferation and invasion through the PI3K/Akt pathway [26]. Also, collagen, a major extracellular matrix protein, is crucial for maintaining tissue structure and facilitating cell growth [27]. Collagen-encoding genes, such as *COL5A2*, have long been considered significant in the prognosis of various cancers [28]. Though less studied than other collagen genes, *COL5A2* has shown overexpression in EC, potentially contributing to tumor progression [29]. Additionally, *TBL1XR1* has been implicated in several gastrointestinal cancers and may promote lymphangiogenesis and lymphatic metastasis in EC via VEGF-C upregulation [30]. Similarly, *TCP1* has emerged lately as a potential prognostic marker for EC [31]. Besides, *CALM2* and *PSMC1*, implicated in various cancers, are known to promote cell proliferation, migration, and survival [32, 33]. Although direct evidence linking *CALM2* and *PSMC1* to EC remains limited, our overall results suggest a potential prognostic role for all 11 key DEGs in EC tumorigenesis.

Moreover, enrichment analysis of the 11 key DEGs further emphasizes their potential involvement in critical processes. Genes were enriched in (GO:0032212), positive regulation of telomere maintenance via telomerase, a fundamental hallmark of cancer that enables limitless replication, requiring inhibition. Similarly, the axon guidance pathway (R-HSA-422475), which tumors may exploit for invasion and metastasis, was significantly enriched. In parallel, enrichment was also observed in pathways related to RNA metabolism (R-HSA-8953854), transport of small molecules (R-HSA-382551), organelle biogenesis and maintenance (R-HSA-1852241), and cellular responses to stimuli (R-HSA-8953897), all of which are essential for normal cellular function but may be co-opted to drive tumorigenesis in state of genomic alterations leading to functional impairment. Dysregulation of mRNA processing, export, and translation has been observed in cancer, where altered RNA metabolism disrupts gene expression and protein synthesis, often mediated by ncRNAs that act as key regulators. These transcriptomic alterations may extensively rewire signaling pathways, influencing both tumor progression and therapeutic response. Likewise, the dysregulation of organelle biogenesis and maintenance has been linked to disruptions in cellular homeostasis, leading to metabolic and structural modifications that can enrich tumor adaptability [34]. Furthermore, the transport of small molecules (R-HSA-382551) enrichment also highlights its role in cellular homeostasis, concerning the uptake of materials and ion homeostasis. However, its aberration is known to drive metabolic adaptation and drug resistance. Notably, *ABCC2*, one of our 11 key DEGs, has been implicated in chemoresistance, where its overexpression favors drug efflux, lowering intracellular drug retention and reducing treatment efficacy [35]. Beyond this, the involvement of our DEGs in the cellular response to stimuli (R-HSA-8953897) related processes has been implicated in promoting mesenchymal stem cell (MSC) differentiation into cancer-associated fibroblasts (CAFs), a process that develops tumor malignancy through biochemical and mechanical cues in the tumor microenvironment.

However, our study is limited by the small cohort size and the lack of *in vitro* functional validation of the identified variants, necessitating further experimental and clinical investigations. Future studies should focus on the functional characterization and therapeutic significance of the identified variants.

This study identified 331 novel missense variants with potential pathogenicity across 274 genes in our cohort. Among these, 23 variants in 11 DEGs exhibited prognostic relevance, structural instability in their corresponding proteins, and involvement in key biological processes driving tumorigenesis, highlighting potential vulnerabilities for precision-targeted therapies.

Conflict of interest

The author has no conflicts with any step of the article preparation.

Consent for publications

The authors have read and approved the final manuscript for publication.

Ethics approval and consent to participate

Human participants were involved in this study, and written informed consent was obtained from all individuals for the use of their data in both the research and subsequent publication. The study protocol was reviewed and approved by the Institutional Review Board of Dow University of Health Sciences (DUHS/Approval/2022/818), in accordance with the Declaration of Helsinki.

Availability of data and material

The Sequencing data supporting the findings of this study have been deposited in the NCBI Sequence Read Archive (SRA) database with Accession ID PRJNA1182118

Author contributions

Conceptualization, Waqas Ahmad Abbasi and Sajida Qureshi; methodology, Waqas Ahmad Abbasi and Muhammad Asif Qureshi; software, Waqas Ahmad Abbasi; validation, Sajida Qureshi, Muhammad Asif Qureshi and Mohammad Saeed Quraishy; formal analysis, Waqas Ahmad Abbasi; data curation, Waqas Ahmad Abbasi, Sajida Qureshi, and Muhammad Asif Qureshi; writing—original draft preparation, Waqas Ahmad Abbasi; writing—review and editing, Sajida Qureshi, Muhammad Asif Qureshi and Mohammad Saeed Quraishy; visualization, Waqas Ahmad Abbasi; funding acquisition, Sajida Qureshi and Muhammad Asif Qureshi

Funding

This research received funding from the Vice Chancellor's Seed Funding initiative of the Dow University of Health Sciences, Karachi, Pakistan.

References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, Jemal A (2024) Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 74 (3): 229-263. d https://oi.org/10.3322/caac.21834
2. AM Filho LM, Ferlay J, Colombet M, Piñeros M, Znaor A (2024) The GLOBOCAN 2022 cancer estimates: Data sources, methods, and a snapshot of the cancer burden worldwide. Int J Cancer. https://doi.org/10.1002/ijc.35278
3. Ferlay J, Ervik M, Lam F, Laversanne M, Colombet M, Mery L, Piñeros M, Znaor A, Soerjomataram I, Bray F (2024) Global cancer observatory: cancer today (version 1.1). Lyon, France: Int Agency Res Cancer. https://gco.iarc.fr/today/en
4. Zhu H, Wang Z, Deng B, Mo M, Wang H, Chen K, Wu H, Ye T, Wang B, Ai D (2023) Epidemiological landscape of esophageal cancer in Asia: Results from GLOBOCAN 2020. Thorac Cancer 14 (11): 992-1003. https://doi.org/10.1111/1759-7714.14835
5. Collins FS, Guyer MS, Chakravarti A (1997) Variations on a theme: cataloging human DNA sequence variation. Science 278 (5343): 1580-1581. https://doi: 10.1126/science.278.5343.1580
6. Qureshi S, Abbasi WA, Jalil HA, Mughal S, Quraishy MS (2024) The First Comprehensive Evaluation of Immuno-Inflammatory Markers for Prognosis in Esophageal Cancer Patients: A South Asian Perspective. Clinics Pract 14 (5): 2071-2079. https://doi.org/10.3390/clinpract14050163
7. Qureshi S, Abbasi WA, Jalil HA, Mughal S, Quraishy MS (2025) Prognostic significance of lymph node ratio in esophageal squamous cell carcinoma: insights from the South Asian population. Front Oncol 14: 1430876. https://doi.org/10.3389/fonc.2024.1430876
8. Li K, Lin Y, Luo Y, Xiong X, Wang L, Durante K, Li J, Zhou F, Guo Y, Chen S (2022) A signature of saliva-derived exosomal small RNAs as predicting biomarker for esophageal carcinoma: a multicenter prospective study. Mol Cancer 21 (1): 21. https://doi.org/10.1186/s12943-022-01499-8
9. Qureshi S, Abbasi WA, Qureshi MA, Jalil HA, Quraishy MS (2024) Identification of PGC as a potential biomarker for progression from barrett's esophagus to esophageal https://doi.org/10.3390/diagnostics14242863
10. Sim N-L, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC (2012) SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Res 40 (W1): W452-W457. https://doi.org/10.1093/nar/gks539
11. Choi Y, Chan AP (2015) PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics 31 (16): 2745-2747. https://doi.org/10.1093/bioinformatics/btv195
12. Chun S, Fay JC (2009) Identification of deleterious mutations within three human genomes. Genome Res 19 (9): 1553-1561. https://doi.org/10.1101/gr.092619.109
13. Schwarz JM, Cooper DN, Schuelke M, Seelow D (2014) MutationTaster2: mutation prediction for the deep-sequencing age. Nat Methods 11 (4): 361-362. https://doi.org/10.1038/nmeth.2890
14. Rogers MF, Shihab HA, Mort M, Cooper DN, Gaunt TR, Campbell C (2018) FATHMM-XF: accurate prediction of pathogenic point mutations via extended features. Bioinformatics 34 (3): 511-513. https://doi.org/10.1093/bioinformatics/btx536
15. Sunyaev S, Ramensky V, Koch I, Lathe III W, Kondrashov AS, Bork P (2001) Prediction of deleterious human alleles. Hum Mol Genet 10 (6): 591-597. https://doi.org/10.1093/hmg/10.6.591
16. Yue P, Moult J (2006) Identification and analysis of deleterious human SNPs. J Mol Biol 356 (5): 1263-1274 https://doi.org/10.1016/j.jmb.2005.12.025
17. Dimmic MW INFERRING SNP FUNCTION USING EVOLUTIONARY. In: Pac Symp Biocomput 2005: Hawaii, USA, 4-8 Jan 2005, 2005. World Scientific, p 382. [No DOI available]
18. Iqbal S, Begum F, Nyamai DW, Jalal N, Shaw P (2023) An integrated computational analysis of high-risk SNPs in angiopoietin-like proteins (ANGPTL3 and ANGPTL8) reveals perturbed protein dynamics associated with cancer. Molecules 28 (12): 4648. https://doi.org/10.3390/molecules28124648
19. Liu X, Hong R, Du P, Yang D, He M, Wu Q, Li L, Wang Y, Chen J, Min Q (2022) The metabolic genomic atlas reveals potential drivers and clinically relevant insights into the etiology of esophageal squamous cell carcinoma. Theranostics 12 (14): 6160. https://doi: 10.7150/thno.70814
20. Donner I, Katainen R, Tanskanen T, Kaasinen E, Aavikko M, Ovaska K, Artama M, Pukkala E, Aaltonen LA (2017) Candidate susceptibility variants for esophageal squamous cell carcinoma. Genes Chromosom Cancer 56 (6): 453-459. https://doi.org/10.1002/gcc.22448
21. Golyan F, Druley T, Abbaszadegan M (2020) Whole-exome sequencing of familial esophageal squamous cell carcinoma identified rare pathogenic variants in new predisposition genes. Clin Transl Oncol 22: 681-693. https://doi.org/10.1007/s12094-019-02174-z
22. Fraser SP, Diss JK, Chioni A-M, Mycielska ME, Pan H, Yamaci RF, Pani F, Siwy Z, Krasowska M, Grzywna Z (2005) Voltage-gated sodium channel expression and potentiation of human breast cancer metastasis. Clin Cancer Res 11 (15): 5381-5389. https://doi.org/10.1158/1078-0432.CCR-05-0327
23. Geuens T, Bouhy D, Timmerman V (2016) The hnRNP family: insights into their role in health and disease. Hum Genet 135: 851-867. https://doi.org/10.1007/s00439-016-1683-5
24. Chen X, Gong R, Wang J, Ma B, Lei K, Ren H, Wang J, Zhao C, Wang L, Yu Q (2022) Identification of HnRNP family as prognostic biomarkers in five major types of gastrointestinal cancer. Curr Gene Ther 22 (5): 449-461. https://doi.org/10.2174/1566523222666220613113647
25. Chen Y, Liu J, Wang W, Xiang L, Wang J, Liu S, Zhou H, Guo Z (2018) High expression of hnRNPA1 promotes cell invasion by inducing EMT in gastric cancer. Oncol Rep 39 (4): 1693-1701. https://doi.org/10.3892/or.2018.6273
26. Fan H, Li J, Jia Y, Wu J, Yuan L, Li M, Wei J, Xu B (2017) Silencing of ribosomal protein L34 (RPL34) inhibits the proliferation and invasion of esophageal cancer cells. Oncol Res 25 (7): 1061. https://doi: 10.3727/096504016X14830466773541

27. Sorushanova A, Delgado LM, Wu Z, Shologu N, Kshirsagar A, Raghunath R, Mullen AM, Bayon Y, Pandit A, Raghunath M (2019) The collagen suprafamily: from biosynthesis to advanced biomaterial development. Adv Mater 31 (1): 1801651. https://doi.org/10.1002/adma.201801651
28. Li J, Wang X, Zheng K, Liu Y, Li J, Wang S, Liu K, Song X, Li N, Xie S (2019) The clinical significance of collagen family gene expression in esophageal squamous cell carcinoma. PeerJ 7: e7705 https://doi.org/10.7717/peerj.7705
29. Li G, Jiang W, Kang Y, Yu X, Zhang C, Feng Y (2020) High expression of collagen 1A2 promotes the proliferation and metastasis of esophageal cancer cells. Ann Transl Med 8 (24): 1672. https://doi.org/10.21037/atm-20-7867
30. Liu F, He Y, Cao Q, Liu N, Zhang W (2016) TBL1XR1 is highly expressed in gastric cancer and predicts poor prognosis. Dis Markers 2016 (1): 2436518. https://doi.org/10.1155/2016/2436518
31. Shao M, Li W, Wang S, Liu Z (2020) Identification of key genes and pathways associated with esophageal squamous cell carcinoma development based on weighted gene correlation network analysis. J Cancer 11 (6): 1393. https://doi.org/10.7150/jca.30699
32. Luo J, Deng L, Zou H, Guo Y, Tong T, Huang M, Ling G, Li P (2023) New insights into the ambivalent role of YAP/TAZ in human cancers. J Exp Clin Cancer Res 42 (1): 130. https://doi.org/10.1186/s13046-023-02704-2
33. Kao T-J, Wu C-C, Phan NN, Liu Y-H, Ta HDK, Anuraga G, Wu Y-F, Lee K-H, Chuang J-Y, Wang C-Y (2021) Prognoses and genomic analyses of proteasome 26S subunit, ATPase (PSMC) family genes in clinical breast cancer. Aging (Albany NY) 13 (14): 17970. https://doi.org/10.18632/aging.203345
34. Graziani V, Crosas-Molist E, George SL, Sanz-Moreno V (2024) Organelle adaptations in response to mechanical forces during tumour dissemination. Curr Opin Cell Biol 88: 102345. https://doi.org/10.1016/j.ceb.2024.102345
35. Alamolhodaei NS, Rashidpour H, Gharaee ME, Behravan J, Mosaffa F (2020) Overexpression of ABCC2 and NF-Kβ/p65 with reduction in cisplatin and 4OH-tamoxifen sensitivity in MCF-7 breast cancer cells: The influence of TNF-α. Pharm Sci 26 (2): 150-158. https//:doi:10.34172/PS.2020.11