Integrated Analysis of mRNA Expression Profiles in Head and Neck Squamous Cell Carcinoma: Insights from Data Mining to Survival Analysis

Sneha Kumari[1], Alvea Tasneem[1], Prithvi Singh[1], Anuska Karmakar[1], Ravins Dohare[1]* Sehrish Imran[2], Rida Rafiqi[3], Neeraj Dohare[4]*, Sunil Kumar[5]*,

Symbiosis School for Online and Digital Learning (SSODL), Symbiosis International (Deemed University), Pune, Maharashtra, India.
[2]Department of Biochemistry, Jamia Hamdard University, New Delhi- 110062, India
[3]Translational Health Science and Technology Institute - Regional Centre for Biotechnology, Faridabad-121001, Haryana, India.
[4]Department of Biochemistry, Daulat Ram College, University of Delhi, New Delhi- 110007, India
[5]Department of Computer Sciences and Information Technology, Central University of Haryana, Mahendragarh-123031, Haryana, India.

*Corresponding Authors:
Prof. Ravins Dohare,Professor,Centre for Interdisciplinary Research in Basic Sciences
Jamia Millia Islamia,New Delhi -110025.India,Tel: +91-9868655958
Email Id: ravinsdohare@gmail.com
Dr. Neeraj Dohare,Assistant Professor ,Department of Biochemistry, Daulat Ram College, University of Delhi, New Delhi- 110007.India,Tel: +91-9654558245
Email Id: n.dohare.150887@gmail.com, neerajdohare@dr.du.ac.in
Dr. Sunil Kumar,Assistant Professor ,Department of Computer Sciences and Information Technology, Central University of Haryana, Mahendragarh-123031, Haryana, India,
Tel: +91 7838123444,Email Id: drsunilk@cuh.ac.in

KEYWORDS
Head and neck squamous cell carcinoma; DEGs; PPIN; GO enrichment; overall survival; mutational analysis.

ABSTRACT:
Background: Head and neck squamous cell carcinoma (HNSCC) is a heterogenous disease involving tumors from the mucosal lining of the oral cavity, pharynx, and larynx, with diverse etiologies, epidemiologies, and treatment strategies. HNSCC carries a poor prognosis, influenced significantly by factors such as HPV status. Innovations in data mining techniques have identified crucial biomarkers for diagnosis and prognosis, facilitating personalized treatment. **Methods:** Present study integrates differential gene expression analysis from the mRNA expression profile dataset of HNSCC downloaded from Gene Expression Omnibus (GEO) database. Screening of differentially expressed genes (DEGs) were performed using R packages. Protein-protein interaction network construction and pathway enrichment analysis revealed significant hub genes which were selected for overall analysis and mutational studies.**Results:** After screening, 322 DEGs in the OSCC tissues were obtained. Construction of protein-protein interaction network (PPIN) led to the selection of hub gene modules, while pathway and gene ontoloy (GO) term enrichment analysis identified interferon alpha/beta signaling pathway as one of the top 10 statistically significant pathways. The Kaplan-Meier (KM) OS and mutational analysis of the selected genes was conducted on patients with HNSCC from The Cancer Genome Atlas (TCGA). Total of 10 candidate genes (IFNA4, LAMA5, HES5, MAP3K5, MAT1A, SLURP1, NCAN, SF3B4, PTPN1, and UQCR10) closely related to the survival rate of patients with HNSCC were identified, and mutational analysis of genes were performed based on TCGA database. SLURP1 showed the highest mutation out of all ten genes in HNSCC patients.

Conclusion: Observed genomic alterations and differential expression of key genes, especially SLURP1, shed light on their potential roles in HNSCC development and patient outcomes, indicating their clinical relevance.

1. Introduction:

Head and Neck Squamous Cell Carcinoma (HNSCC) poses a significant challenge as the sixth most common cancer worldwide, affecting more than 600,000 new patients each year. It commonly occurs in the upper aerodigestive tract, including the oral cavity, pharynx, and larynx HNSCC can originate from various sites such as lips, tongue, nasopharynx, oropharynx, or hypopharynx, making it a heterogeneous disease (Rothenberg and Ellisen, 2012). The oncology domain is characterized by the malignant transformation of mucosal epithelium within the oral cavity, pharynx, and larynx (Johnson et al., 2020). The perplexing malignancy signals us to look into a complex exploration that resolves its intricate epidemiological variation, etiological factors, immunological complexities, and the evolving landscape of therapeutic strategies.

HNSCC arises from a complex interaction of pathogens (causative agents), each contributing to the complexity of this disease. Amongst all agents, tobacco, a reservoir of carcinogens, becomes the petrifying protagonist. The carcinogens settled within tobacco smoke penetrate the delicate mucosal linings of the head and neck, setting off a cascade of cellular events that lays the foundation for malignancy. Alongside tobacco is another protagonist, alcohol consumption, which collaborates with tobacco to intensify the risk of HNSCC. The slick amalgamation of tobacco and alcohol puts together a threatening web of genetic mutations, tumor initiation and DNA damage (Chamoli et al., 2021; Rivera, 2015; Urashima et al., 2013). Patients with HNSCC are often diagnosed at an advanced stage, typically without any prior signs or pre-cancerous lesions. Major symptoms involves pain or difficulty swallowing, weight loss, lump in throat or mouth, persistent sore throat, and hoarseness. Diagnosis typically involves a combination of physical examination, imaging studies, and biopsy (Chow, 2020).

The molecular landscape of HNSCC is complex, with various genetic alterations influencing tumor behavior and patient prognosis. Key genetic markers include mutations in the *TP53* gene, amplification of the *EGFR* gene, and human papillomavirus (HPV) integration in a subset of tumors (Leemans et al., 2018). Therapeutic options are generally based on the stage of malignancy and tumor location, and may include surgery, radiation therapy, chemotherapy, or targeted therapy. The prognosis varies widely based on factors such as tumor stage, HPV status, and overall health of the patient (Argiris et al., 2008). HPV-positive HNSCC typically has a better prognosis (Hübbers and Akgül, 2015; Sabatini and Chiocca, 2020). Multimodal approaches and multidisciplinary care are often necessary for managing HNSCC. Certain FDA-approved drugs like cetuximab and immune checkpoint inhibitors show promise in treating recurrent or metastatic cases (Goel et al., 2022; Poulose and Kainickal, 2022). Molecular and immune profiling offers the potential for identifying biomarkers that can enhance targeted therapies and improve patient outcomes. This Primer provides insights into different types of HNSCC, their causes, and how understanding their origin shapes treatment strategies.

Innovative research in HNSCC has shed light on the critical role of mRNA expression profiles. A pivotal study exhibited robust rank aggregation in deciphering complex gene co-expression networks in HNSCC, marking a significant stride in molecular oncology (Cao et al., 2020). Consequently, another insightful research piece examines gene expression alterations in HNSCC under PD1/PD-L1 inhibitor treatment, steering towards more targeted therapeutic approaches (Foy et al., 2022). Further, an investigation reveals how different matrices influence gene expression in HNSCC cells, a crucial factor in optimizing in vitro therapy testing (Hyytiäinen et al., 2023).

Meanwhile, a study provides a comparative analysis of expression profiles between HNSCC tumors and normal tissues, uncovering key molecular distinctions (Lemaire et al., 2003). Additionally, groundbreaking work employs single-cell RNA sequencing to unravel the intricate gene expression patterns unique to HNSCC, offering a more nuanced understanding of its molecular dynamics (Yu et al., 2020). Collectively, these studies propel forward our comprehension of HNSCC at the molecular level, opening new avenues for diagnosis and treatment.

Furthering the studies in the field of HNSCC has leveraged data mining techniques to unearth critical biomarkers for diagnosis and prognosis. For instance, innovative research highlighted the importance of survival analysis in identifying these biomarkers, which could revolutionize patient treatment strategies (Ju et al., 2023). Furthermore, an investigation reveals how data mining can isolate specific transcription factors that predict HNSCC outcomes, paving the way for tailored therapies (Zhang et al., 2019). Additionally, a study delves into the The Cancer Genome Atlas (TCGA) database to pinpoint genes within the HNSCC microenvironment, linking them to patient survival prospects (Ran et al., 2021). A similar exploration identifies six proteins viz. ERALPHA, HER3, BRAF, P27, RAPTOR, and E2F1 are novel prognostic markers the offer potential targets for future treatments (Wu et al., 2020). Lastly, an integrative approach combines mRNA expression data from diverse sources (viz. GEO and TCGA, etc.), employing bioinformatics for comprehensive survival analysis, thereby enriching our molecular understanding of HNSCC (Shen et al., 2019). Ongoing research in HNSCC focuses on understanding the molecular drivers of the disease, developing targeted therapies, and improving detection and treatment strategies for HPV-associated cancers (Gillison et al., 2015). HNSCC presents a significant challenge in oncology, marked by its heterogeneity and variable patient outcomes. This study embarks on an integrated analysis of mRNA expression profiles in HNSCC, utilizing data mining techniques to draw a correlation with patient survival, a method inspired by recent advancements in genomic profiling.

By analyzing a comprehensive dataset of mRNA expressions from HNSCC patients, we applied advanced data mining algorithms to identify key patterns and biomarkers. This approach aligns with the methodologies suggested in (Zoabi and Shomron, 2021), enabling the discovery of specific mRNA signatures correlated with clinical outcomes, including treatment response and survival rates.

2. Materials and Methods

2.1 HNSCC mRNA data extraction and DEA

The mRNA expression profiling datasets for HNSCC were downloaded from NCBI-GEO (Clough and Barrett, 2016) (https://www.ncbi.nlm.nih.gov/geo/) using specific keywords such as 'HNSCC' or 'Head and Neck Squamous Cell Carcinoma.' The criteria for selecting suitable datasets included: (i) expression profiling with a sample count greater than 25, (ii) organism set to Homo sapiens, (iii) dataset publication date within the last 10 years, and (iv) datasets containing both control and tumor patient samples. Datasets involving non-human samples, case reports, review articles, abstracts, and cell-line-based experimental designs were excluded. Series matrix expression file of the chosen dataset was downloaded. Batch Correction was performed using ARSyNseq function available within NOISeq (v 2.16) R package (Tarazona et al., 2012) and mapping of probe IDs was performed using feature data stored in the expression set object of dataset. A two-sample statistical t-test was utilized for calculation of the p-value of each gene between control and tumor samples followed by obtaining their $|log_2(fold\ change)| > 1.5$ benjamini-hochberg (BH) p-values utilizing the Limma R package (Ritchie et al., 2015). The differentially expressed genes (DEGs)

corresponding to Benjamini-Hochberg (BH) corrected $p - value < 0.05$ were bifurcated as up- and down regulating, respectively.

2.2 *Construction of HNSCC-associated protein-protein interaction (PPI) network*

For the identificaiton of hub proteins associated with HNSCC-associated DEGs, protein-protein interaction network (PPIN) was constructed using Search Tool for the Retrieval of Interacting Genes (STRING) v12.0 database (https://string-db.org/) (Szklarczyk et al., 2021). Utilizing all active interaction sources ─ experimental data, text mining, predicted databases, neighborhood, co-expression, fusion, and co-occurrence ─ the DEGs-associted proteins were analyzed for their interactions with neighboring proteins at a highest confidence level (interaction score ≥ 0.90). The PPIN was further visualized and analyzed using Cytoscape v 3.10 (Shannon et al., 2003).

2.3 Pathway and GO term enrichment analyses of key HNSCC-associated genes

For pathway and gene ontology (GO) term enrichment analysis, EnrichR (https://maayanlab.cloud/Enrichr/) was accessed (Chen et al., 2013). Biological pathway libraries such as Reactome 2024, along with GO terms (Biological Process, Molecular Function and Cellular Compartment) available within EnrichR, were employed to identify the top 10 significantly enriched pathway and GO terms ($p - value < 0.05$).

2.4 Overall survival (OS) and mutational analyses

Kaplan-Meier (KM) plotter (https://kmplot.com/analysis/) (Lánczky and Győrffy, 2021) was accessed to compile OS data corresponding to PPIN-associated elements. The parameters used for drawing KM plots are mentioned previously (Singh et al., 2023). Using cBioPortal (https://www.cbioportal.org/) prognostically significant hub genes were studied (Gao et al., 2013). The Head and Neck Squamous Cell Carcinoma (TCGA, Firehose legacy) dataset was chosen, and the 530 tumor samples present in the dataset were selected for mutational analysis (https://www.cbioportal.org/study/summary?id=hnsc_tcga).

3. Results:

3.1 HNSCC mRNA data extraction and differentially expressed genes analysis

Based on the inclusion and exclusion criteria for selecting mRNA expression profiles, the HNSCC dataset with accession number GSE83519 was obtained from NCBI-GEO. This dataset included 44 PBMC patient samples (22 tumor and 22 normal samples). Pre-processing and normalization of the raw expression data was performed, and batch effect correction was applied to remove batch variables. Probe IDs were mapped to their corresponding genes ad duplicate gene identifiers were removed by averging the expression values of the duplicates, resulting in 322 unique gene symbols.

Differentially expressed genes (DEGs) analysis was conducted for the two groups (22 tumor samples paired with 22 normal mucosa samples) using a paired *t-test* in the limma R package. A total of 322 genes were identified as differentially expressed based on $BH - adjusted\ p - value < 0.05$ and $|log_2(\ Fold\ Change)| > 1.5$ between the groups. The DEGs consisted of 207 upregulated and 115 downregulated genes. The volcano plot (**Figure 1)** illustrate the distribution of significant (coloured plots) as well as non-significant genes (black points) in the group. **Table 1** lists the top 10 upregulated and top 10 downregulated DEGs for HNSCC. The most highly upregulated and downregulated DEGs were *DRG1* $[log_2(\ Fold\ Change) = 6.563]$ and *B4GALT4* = $[log_2(Fold\ Change) = -6.323]$, respectively.

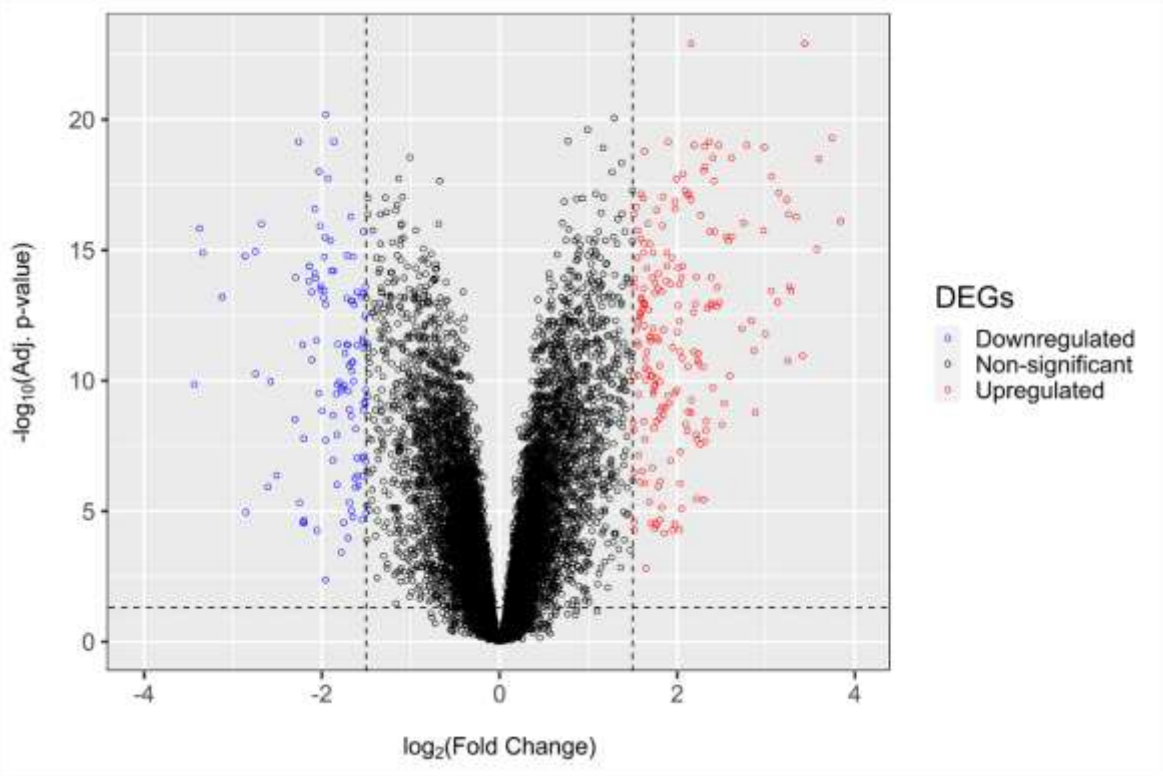


Figure 1: Volcano plots show the distribution of significant (coloured plots) as well as nonsignificant (black-coloured points) genes in the HNSCC tumor and normal groups.

Table 1: List of top 10 up and downregulated differentially expressed genes among HNSCC patients and control samples.

Upregulated Genes	$log_2(FC)$	**P-value**	**Adjusted P-value**
DRG1	6.562734	9.25E-19	1.91E-16
ZNF285	4.634537	1.02E-15	6.87E-14
ATP6V0A4	3.839575	3.07E-19	7.87E-17
TEX10	3.743346	1.56E-23	4.98E-20
MBD3	3.597593	3.76E-22	3.20E-19
RNASEH1-AS1	3.571157	5.97E-18	9.40E-16
TRIM60	3.435328	1.28E-27	1.23E-23
TMPRSS11E	3.410732	3.66E-13	1.11E-11
TAAR6	3.341326	1.99E-19	5.27E-17
CEBPD	3.282793	4.61E-16	3.63E-14
Downregulated Genes	$log_2(FC)$	**P-value**	**Adjusted P-value**
B4GALT4	−6.32342	9.16E-20	2.89E-17
EHMT1-IT1	−6.24032	8.13E-20	2.65E-17
FAM83E	−6.16861	1.39E-19	4.06E-17
ZNF639	−4.73405	3.83E-22	3.20E-19
PARVB	−3.43512	6.35E-12	1.38E-10
CD207	−3.37588	6.57E-19	1.48E-16
MIR503HG	−3.33775	8.50E-18	1.26E-15
SCG2	−3.12126	9.08E-16	6.34E-14
LRRC7	−2.86418	1.21E-17	1.68E-15
MAGEA2B	−2.85604	2.07E-06	1.11E-05

Abbreviations: $\boldsymbol{log2}$ ($\boldsymbol{FC}$) – Logarithmic Fold Change

3.2 Construction of HNSCC-associated PPIN, pathway enrichment, and GO-term analysis

For the construction of HNSCC-associated PPI network, all 322 DEGs were submitted to the STRING database, and the network was generated using the highest confidence score of > 0.9. The resulting PPI network consisted of 34 nodes and 23 edges, with 17 upregulated genes (colored in pink) and 17 were downregulated (colored in seagreen) **(Figure 2)**.

The DEGs involved in the highest-confidence PPI network were used for collecting pathway and GO term enrichment data utilizing various libraries (Reactome 2024 for pathway, and Human Phenotype Ontologies 2023 for GO term) within EnrichR database. The top 10 pathways and GO terms corresponding to a $p-value < 0.05$ were regarded as statistically significant (**Tables 2-5**). A total of 17, 14, 16, and 6 DEGs participated in top 10 signifiant pathway, GO-BP, MF, CC terms, respectively. Most significant pathway, GO-BP, GO-MF, and GO-CC terms were Interferon alpha/beta signaling pathway ($p-value = 9.08 \times 10^{-06}$), type I interferon-mediated signaling pathway ($p-value = 2.36 \times 10^{-05}$), acetylcholine receptor regulator activity ($p-value = 2.52 \times 10^{-04}$), and death-inducing signaling complex ($p-value = 5.86 \times 10^{-05}$).

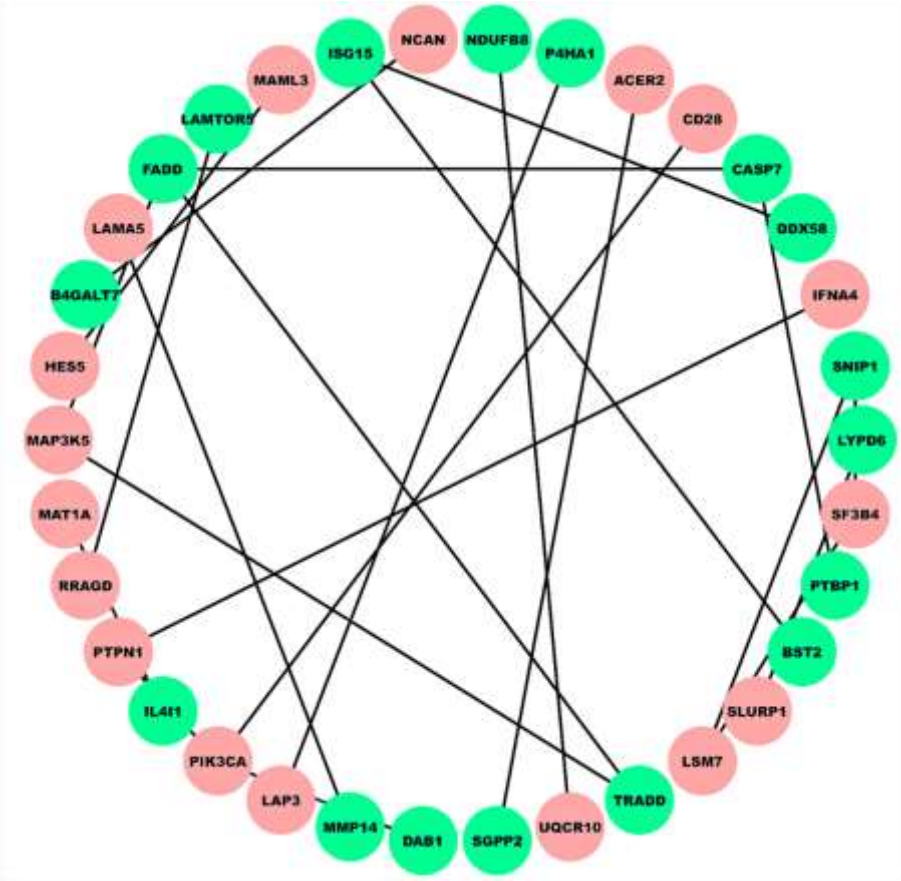


Figure 2: PPI network comprising 34 nodes and 23 edges corresponding to STRING $interaction\ score > 0.9$. Pink and seagreen node colors signify up and downregulated genes, respectively.

Table 2: List of top 10 significant pathways from Reactome 2024, their p-value, and gene names.

Pathways	$p - value$	**Gene Names**
Interferon Alpha Beta Signaling	9.08E-06	BST2; IFNA4; PTPN1; ISG15
CASP8 Activity Is Inhibited	1.53E-04	TRADD; FADD
Regulation by c-FLIP	1.53E-04	TRADD; FADD
Dimerization of Procaspase-8	1.53E-04	TRADD; FADD
NOTCH2 Intracellular Domain Regulates Transcription	1.83E-04	MAML3; HES5
Sphingolipid Catabolism	1.83E-04	ACER2; SGPP2
Defective RIPK1-mediated Regulated Necrosis	1.83E-04	TRADD; FADD
Signaling by MET	3.25E-04	LAMA5; PTPN1; PIK3CA
mRNA Splicing - Major Pathway	3.83E-04	SF3B4; PTBP1; LSM7; SNIP1
DDX58 IFIH1-mediated Induction of Interferon-Alpha Beta	4.17E-04	IFNA4; ISG15; FADD

Table 3: List of top 10 significant GO-BP terms, their p-value, and gene names.

GO-BP	$p - value$	**Gene Names**
Regulation Of Type I Interferon-Mediated Signaling Pathway (GO:0060338)	2.36E-05	PTPN1; ISG15; FADD
Positive Regulation Of TOR Signaling (GO:0032008)	6.92E-05	PIK3CA; RRAGD; LAMTOR5
Regulation Of Viral Genome Replication (GO:0045069)	2.00E-04	BST2; CD28; ISG15
Positive Regulation Of I-kappaB kinase/NF-kappaB Signaling (GO:0043123)	2.14E-04	BST2; TRADD; FADD; LAMTOR5
Positive Regulation Of Intracellular Signal Transduction (GO:1902533)	2.29E-04	BST2; PIK3CA; TRADD; RRAGD; FADD; LAMTOR5
Regulation Of TOR Signaling (GO:0032006)	3.13E-04	PIK3CA; RRAGD; LAMTOR5
Regulation Of Adaptive Immune Response (GO:0002819)	4.22E-04	IL4I1; FADD
Sphingosine Metabolic Process (GO:0006670)	4.71E-04	ACER2; SGPP2
Regulation Of B Cell Differentiation (GO:0045577)	4.71E-04	IL4I1; MMP14
Apoptotic Process (GO:0006915)	5.83E-04	CASP7; PIK3CA; TRADD; FADD

Table 4: List of top 10 significant GO-MF terms, their p-value, and gene names.

GO-MF	$p-value$	Gene Names
Acetylcholine Receptor Regulator Activity (GO:0030548)	2.52E-04	LYPD6; SLURP1
Oxidoreduction-Driven Active Transmembrane Transporter Activity (GO:0015453)	0.003797894	NDUFB8; UQCR10
Receptor Tyrosine Kinase Binding (GO:0030971)	0.004516979	PTPN1; TRADD
N-acylsphingosine Amidohydrolase Activity (GO:0017040)	0.0084719	ACER2
Cysteine-Type Endopeptidase Activity Involved In Execution Phase Of Apoptosis (GO:0097200)	0.0084719	CASP7
Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Activity (GO:0046934)	0.0084719	PIK3CA
Acetylcholine Receptor Inhibitor Activity (GO:0030550)	0.0084719	LYPD6
RNA Binding (GO:0003723)	0.008542296	SF3B4; BST2; PTBP1; PTPN1; CASP7; LSM7; SNIP1
Primary Amine Oxidase Activity (GO:0008131)	0.010157909	IL4I1
Sphingosine-1-Phosphate Phosphatase Activity (GO:0042392)	0.010157909	SGPP2

Table 5: List of top 10 significant GO-CC terms, their p-value, and gene names.

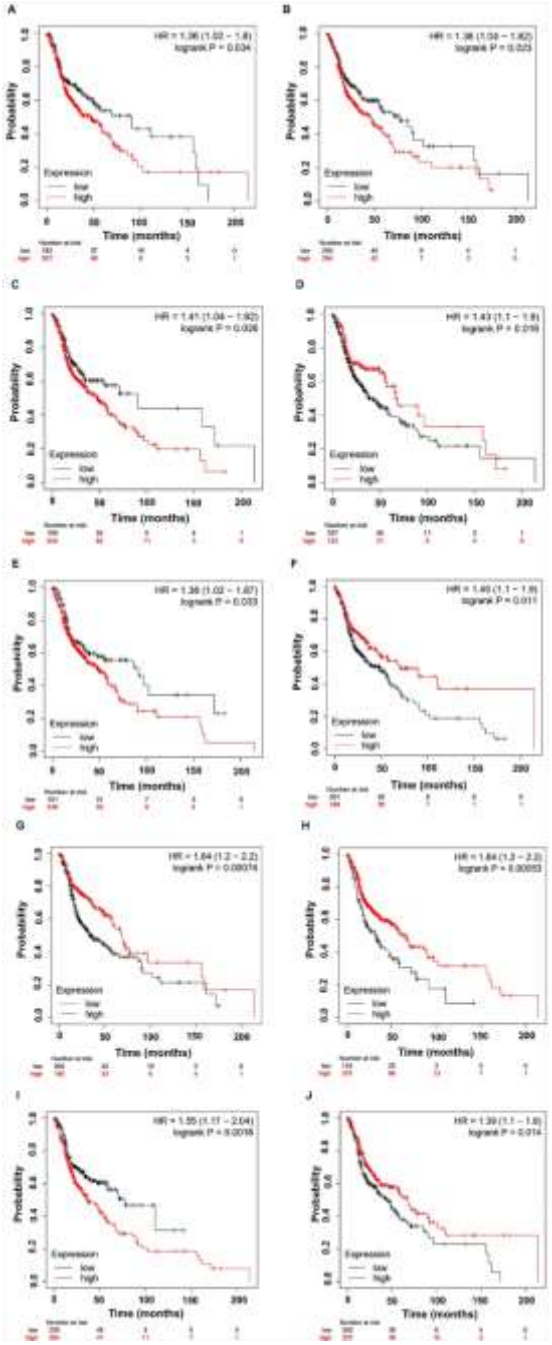
GO-CC	$p-value$	Gene Names
Death-Inducing Signaling Complex (GO:0031264)	5.86E-05	TRADD; FADD
U2-type Precatalytic Spliceosome (GO:0071005)	7.84E-05	SF3B4; LSM7; SNIP1
Precatalytic Spliceosome (GO:0071011)	9.37E-05	SF3B4; LSM7; SNIP1
Spliceosomal snRNP Complex (GO:0097525)	1.23E-04	SF3B4; LSM7; SNIP1
U2-type Spliceosomal Complex (GO:0005684)	4.77E-04	SF3B4; LSM7; SNIP1
U2 snRNP (GO:0005686)	8.21E-04	SF3B4; SNIP1
U12-type Spliceosomal Complex (GO:0005689)	0.001031333	SF3B4; LSM7
Macropinosome (GO:0044354)	0.010157909	MMP14
Pinosome (GO:0044352)	0.010157909	MMP14
U6 snRNP (GO:0005688)	0.011841135	LSM7

3.2 OS and mutational analyses

The overall survival of patients depends on two key factors: time and event, collectively known as as"time to event". Time refers to the duration until the event occurs after exposure (e.g. diagnosis), and the event can be death. Survival time is of particular interest; as it helps determine how long a patient survives based on the expression of specific genes. In this study, 24 genes were identified as negaitvely impacting survival, while 10 genes were involve in survival analysis (**Figure 3**). A Kalan-Meier (KM) plot was generated from the survival analysis, depicting the relationship between probability and time. The plot compares survival times under two conditions: high and

low gene expression, providing the log p-value of each gene (**Table 6**). From the DEGs file, we can determine whether the particular gene is upregulated or downregulated. If the $log\ p - value$ of gene is is less than 0.05, it is considered as significant; if it exceed 0.05, it is considered as non-significant. All prognostically significant hub genes (**Figure 4, Table 7**) were found to be mutated in 139 (26%) patient samples. Among these, the *SLURP1* gene showed alteration (or amplification) in 56 cases (10.57%).

Figure 3: Association of different genes with increased overall survival of HNSCC patients across ten independent cohorts. Kaplan-Meier plots of patient survival stratified by different gene expressions. (A) IFNA4 (B) LAMA5 (C) HES5 (D) MAP3K5 (E) MAT1A (F) SLURP1 (G) NCAN (H) SF3B4 (I) PTPN1 (J) UQCR10. Red and black color signify higher and lower expression groups.



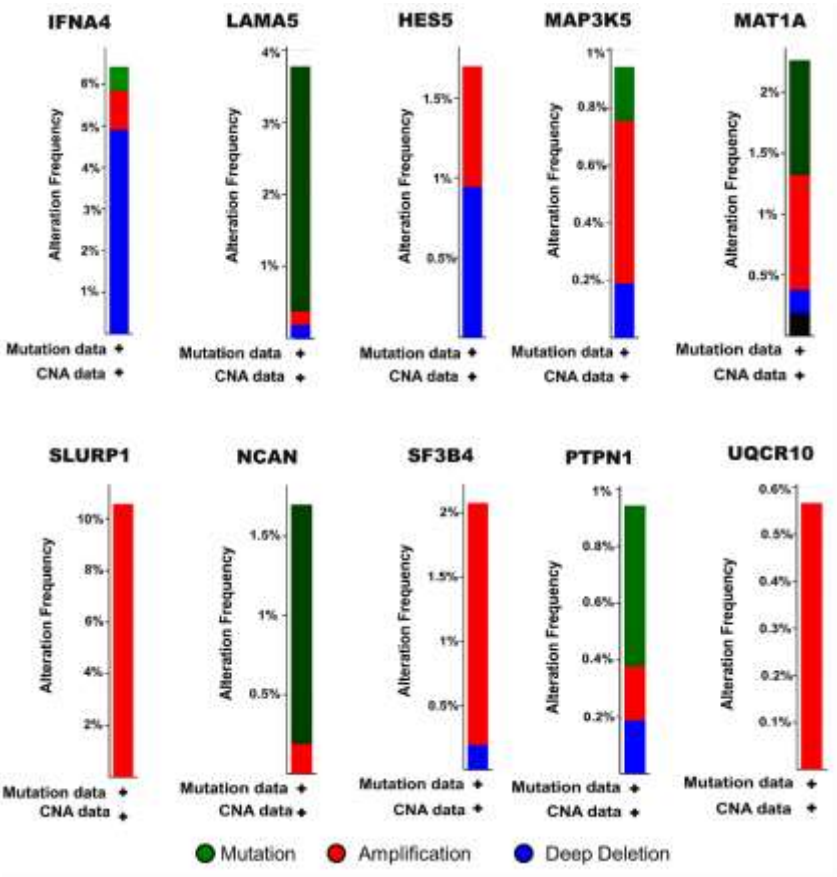


Figure 3: Alteration frequency percentage of IFNA4, LAMA5, HES5, MAP3K5, MAT1A, SLURP1, NCAN, SF3B4, PTPN1, and UQCR10.

Table 6. List of top prognostically significant genes, and their median survival times (high and low expression cohort).

Items	Low Expression Cohort (Months)	High Expression Cohort (Months)
*UQCR*10	77.3	46.47
*HES*5	32.67	65.73
*IFNA*4	77.3	36.43
*LAMA*5	46.6	66.73
*MAP*3*K*5	33.1	68.8
*MAT*1*A*	91.37	48.87
NCAN	47.67	69.43
*PTPN*1	90.1	46.6
*SF*3*B*4	90.57	48.63
*SLURP*1	40.07	66.73

Table 7. Types of mutations and their frequencies for all prognostically significant genes.

	Types Of Mutations		
Gene	**Amplification**	**Deep Deletion**	**Mutation**
IFNA4	0.94%	4.91%	0.57%
LAMA5	0.19%	0.19%	3.4%
HES5	0.75%	0.94%	–
MAP3K5	0.94%	0.19%	0.94%
MAT1A	0.57%	0.19%	0.19%
SLURP1	10.57%	–	–
NCAN	0.19%	–	1.51%
SF3B4	1.89%	0.19%	–
PTPN1	0.19%	0.19%	0.57%
UQCR10	0.57%	–	–

4. Discussion

Head and neck tumor include oncogenic development of the mucosal epithelium in the oral cavity, pharynx, and larynx. The most common causative agents include smoke, alcohol consumption, tobacco containing carcinogen, exposure to harmful rays (such as ionizing and non-ionizing radiations).

Occurrence of HNSCC may differ depending upon the exposure to carcinogenic agents, family history and geographical areas. The studies have observed no significant increase and decrease in the incidence of HNSCC among the young people and no significant difference in cases of HNSCC among different age group of people, and cases are observed more in men.

This research encompasses key findings (viz. Data Extraction and Differential Expression Analysis, Protein-Protein Interaction Network and Functional Enrichment Analyses, Survival and Mutational Analyses), clinical and therapeutic implications, the strengths and limitations, and future research direction.

Identification of new therapeutic targets and development of therapeutic agents in head and neck tumor are strongly being followed. Computational approaches play a vital role in achieving this as it significantly speeds up the identification of therapeutic targets and the development of therapeutic molecules in drug discovery. Therefore, different therapeutic drugs can be proposed using computational tools by observing the PPI network and the survival analysis of the genes (Mody et al., 2021; Tutar, 2014). The exploration of HNSCC through mRNA expression profiling has revolutionized our understanding of this complex disease. The integration of genomic data and its comprehensive analysis has provided a detailed landscape of the genetic alterations in HNSCC (Cao et al., 2020). This backdrop of advanced genomic understanding sets the stage for identifying specific biomarkers and pathways critical to the disease's progression and patient outcomes.

In the realm of specific biomarkers, chromosomal aberrations in HNSCC have garnered significant attention. For instance, the amplification of chromosome 3q is linked to the overexpression of key oncogenes like *PIK3CA*, contributing to tumor progression (Lemaire et al., 2003). Similarly, deletions in chromosome 9p21, impacting tumor suppressor genes such as *CDKN2A*, have been associated with poor prognosis in HNSCC (Yu et al., 2020). These chromosomal changes not only offer insights into the molecular mechanisms driving HNSCC but also present potential targets for novel therapeutic interventions.

The utilization of GEO terms and pathways has furthered our understanding of HNSCC at a molecular level. GEO datasets have been quite instrumental to identify deregulated pathways in HNSCC, such as the MAPK signaling pathway, known for its role in cell proliferation and differentiation (Hyytiäinen et al., 2023). GO serves as a powerful framework for understanding the molecular mechanisms underlying HNSCC pathogenesis. This paper explores the multifaceted utility of GO in elucidating the biological processes, molecular functions, and cellular components implicated in HNSCC without redundancy. It delineates how GO annotations aid in deciphering the intricacies of HNSCC etiology, progression, and potential therapeutic Another critical pathway, the JAK-STAT signaling pathway, implicated in immune response and tumor microenvironment interactions, has been identified in HNSCC research (Shen et al., 2019). These pathways not only provide a deeper understanding of HNSCC pathophysiology but also open up new avenues for targeted therapy.

5. Conclusion

Our study suggests that the integrated analysis of mRNA expression profiles in HNSCC, backed by genomic data and advanced analytical techniques, has significantly advanced our understanding of this disease. The identification of chromosomal aberrations and the elucidation of key signaling pathways offer a more nuanced view of HNSCC's molecular landscape. Understanding the genetic landscape of HNSCC can also help identify mechanisms of treatment resistance and relapse. Genes that are highly expressed or mutated in treatment-resistant cases can be targets for developing new drugs or combination therapies. The pathways and biological processes identified in our research can find new therapeutic targets. For example, if a particular pathway is found to be significantly involved in the progression of HNSCC, drugs that inhibit this pathway could be developed or repurposed for treatment.

The findings suggest potential molecular mechanisms underlying HNSCC progression. Identified hub genes and pathways associated with immune response regulation, morphogenesis, and specific cellular functions may serve as therapeutic targets or prognostic indicators. Additionally, the observed alterations and mutations in key genes, especially *SLURP1*, shed light on their potential roles in HNSCC development and patient outcomes, indicating their clinical relevance.

6. Abbreviations:

HNSCC: Head and Neck Squamous Cell Carcinoma
NCBI: National Center for Biotechnology Information
GEO: Genome Expression Omnibus
BH: Benjamini- Hochberg
DEGs: Differentially Expressed Genes
PPI: Protein-Protein Interaction
GO: Gene Ontology
BP: Biological Process
MF: Molecular Function
CC: Cellular Compartment
KM: Kaplan- Meier
OS: Overall Survival

7. Statements and Declaration

Acknowledgements: R.D would like to thank the Anusandhan National Research Foundation (ANRF), Department of Science and Technology, Govt. of India [G.No. EEQ/2023/00980].
Conflicts of Interest: The authors declare no conflict of interest.

Funding: No funding.
Supplementary Information: Supplementary data of this article is provided in the supplementary material.

Author's Contributions:

Conceptualization: Neeraj Dohare, Sunil Kumar, and Ravins Dohare

Methodology: Sneha Kumari, Alvea Tasneem, Prithvi Singh, Anuska Karmakar, Sehrish Imran, and Rida Rafiqi

Investigation: Sneha Kumari, Anuska Karmakar, Sehrish Imran, and Rida Rafiqi

Software: Anuska Karmakar, Sehrish Imran, and Rida Rafiqi

Formal Analysis: Prithvi Singh, Anuska Karmakar, Sehrish Imran, and Rida Rafiqi

Data Curation: Anuska Karmakar, Sehrish Imran, and Rida Rafiqi

Validation: Sneha Kumari, Alvea Tasneem, and Prithvi Singh

Visualization: Alvea Tasneem, Anuska Karmakar, Sehrish Imran, and Rida Rafiqi

Writing- Original Draft: Sneha Kumari, Alvea Tasneem, Anuska Karmakar, Sehrish Imran, and Rida Rafiqi

Writing- Review & Editing: Sneha Kumari, Alvea Tasneem, Prithvi Singh, Neeraj Dohare, Sunil Kumar, and Ravins Dohare

Supervision: Neeraj Dohare, Sunil Kumar, and Ravins Dohare

Project administration: Neeraj Dohare, Sunil Kumar, and Ravins Dohare

Data Availability Statement:

The dataset used in our study was downloaded from National Center for Biotechnology Information-Gene Expression Omnibus under accession number GSE83519 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE83519). The publicly available data in Figure 3 and Table 6 (TCGA, Firehose Legacy) are available in the cBioPortal database (http://www.cbioportal.org/study/summary?id=hnsc_tcga) and are used to perform mutational analysis. The publicly available data in Figure 4 and Table 5 are available in the KM-Plotter pan-Cancer database (https://kmplot.com/analysis/).

8. References:

Argiris, A., Karamouzis, M.V., Raben, D., Ferris, R.L., 2008. Head and neck cancer. The Lancet 371, 1695–1709. https://doi.org/10.1016/S0140-6736(08)60728-X

Cao, Z., Ao, Y., Guo, Y., Zhou, S., 2020. Comprehensive Analysis of mRNA Expression Profiles in Head and Neck Cancer by Using Robust Rank Aggregation and Weighted Gene Coexpression Network Analysis. BioMed Research International 2020, 1–21. https://doi.org/10.1155/2020/4908427

Chamoli, A., Gosavi, A.S., Shirwadkar, U.P., Wangdale, K.V., Behera, S.K., Kurrey, N.K., Kalia, K., Mandoli, A., 2021. Overview of oral cavity squamous cell carcinoma: Risk factors, mechanisms, and diagnostics. Oral Oncol 121, 105451. https://doi.org/10.1016/j.oraloncology.2021.105451

Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., Ma'ayan, A., 2013. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics 14, 128. https://doi.org/10.1186/1471-2105-14-128

Chow, L.Q.M., 2020. Head and Neck Cancer. N Engl J Med 382, 60–72. https://doi.org/10.1056/NEJMra1715715

Clough, E., Barrett, T., 2016. The Gene Expression Omnibus Database, in: Mathé, E., Davis, S. (Eds.), Statistical Genomics, Methods in Molecular Biology. Springer New York, New York, NY, pp. 93–110. https://doi.org/10.1007/978-1-4939-3578-9_5

Foy, J.-P., Karabajakian, A., Ortiz-Cuaran, S., Boussageon, M., Michon, L., Bouaoud, J., Fekiri, D., Robert, M., Baffert, K.-A., Hervé, G., Quilhot, P., Attignon, V., Girod, A., Chaine, A., Benassarou, M., Zrounba, P., Caux, C., Ghiringhelli, F., Lantuejoul, S., Crozes, C., Brochériou, I., Pérol, M., Fayette, J., Bertolus, C., Saintigny, P., 2022. Datasets for gene expression profiles of head and neck squamous cell carcinoma and lung cancer treated or not by PD1/PD-L1 inhibitors. Data in Brief 44, 108556. https://doi.org/10.1016/j.dib.2022.108556

Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., Cerami, E., Sander, C., Schultz, N., 2013. Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. Sci. Signal. 6. https://doi.org/10.1126/scisignal.2004088

Gillison, M.L., Chaturvedi, A.K., Anderson, W.F., Fakhry, C., 2015. Epidemiology of Human Papillomavirus–Positive Head and Neck Squamous Cell Carcinoma. JCO 33, 3235–3242. https://doi.org/10.1200/JCO.2015.61.6995

Goel, B., Tiwari, A.K., Pandey, R.K., Singh, A.P., Kumar, S., Sinha, A., Jain, S.K., Khattri, A., 2022. Therapeutic approaches for the treatment of head and neck squamous cell carcinoma–An update on clinical trials. Translational Oncology 21, 101426. https://doi.org/10.1016/j.tranon.2022.101426

Hübbers, C.U., Akgül, B., 2015. HPV and cancer of the oral cavity. Virulence 6, 244–248. https://doi.org/10.1080/21505594.2014.999570

Hyytiäinen, A., Korelin, K., Toriseva, M., Wilkman, T., Kainulainen, S., Mesimäki, K., Routila, J., Ventelä, S., Irjala, H., Nees, M., Al-Samadi, A., Salo, T., 2023. The effect of matrices on the gene expression profile of patient-derived head and neck carcinoma cells for in vitro therapy testing. Cancer Cell Int 23, 147. https://doi.org/10.1186/s12935-023-02982-y

Johnson, D.E., Burtness, B., Leemans, C.R., Lui, V.W.Y., Bauman, J.E., Grandis, J.R., 2020. Head and neck squamous cell carcinoma. Nat Rev Dis Primers 6, 92. https://doi.org/10.1038/s41572-020-00224-3

Ju, G., Yao, Z., Zhao, Y., Zhao, X., Liu, F., 2023. Data mining on identifying diagnosis and prognosis biomarkers in head and neck squamous carcinoma. Sci Rep 13, 10020. https://doi.org/10.1038/s41598-023-37216-8

Lánczky, A., Győrffy, B., 2021. Web-Based Survival Analysis Tool Tailored for Medical Research (KMplot): Development and Implementation. J Med Internet Res 23, e27633. https://doi.org/10.2196/27633

Leemans, C.R., Snijders, P.J.F., Brakenhoff, R.H., 2018. The molecular landscape of head and neck cancer. Nat Rev Cancer 18, 269–282. https://doi.org/10.1038/nrc.2018.11

Lemaire, F., Millon, R., Young, J., Cromer, A., Wasylyk, C., Schultz, I., Muller, D., Marchal, P., Zhao, C., Melle, D., Bracco, L., Abecassis, J., Wasylyk, B., 2003. Differential expression profiling of head and neck squamous cell carcinoma (HNSCC). Br J Cancer 89, 1940–1949. https://doi.org/10.1038/sj.bjc.6601373

Mody, M.D., Rocco, J.W., Yom, S.S., Haddad, R.I., Saba, N.F., 2021. Head and neck cancer. Lancet 398, 2289–2299. https://doi.org/10.1016/S0140-6736(21)01550-6

Poulose, J.V., Kainickal, C.T., 2022. Immune checkpoint inhibitors in head and neck squamous cell carcinoma: A systematic review of phase-3 clinical trials. WJCO 13, 388–411. https://doi.org/10.5306/wjco.v13.i5.388

Ran, Q.-C., Long, S.-R., Ye, Y., Xie, C., XuXiao, Z.-L., Liu, Y.-S., Pang, H.-X., Sunchuri, D., Teng, N.-C., Guo, Z.-L., 2021. Mining TCGA database for prognostic genes in head and neck squamous cell carcinoma microenvironment. Journal of Dental Sciences 16, 661–667. https://doi.org/10.1016/j.jds.2020.09.017

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., Smyth, G.K., 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research 43, e47–e47. https://doi.org/10.1093/nar/gkv007

Rivera, C., 2015. Essentials of oral cancer. Int J Clin Exp Pathol 8, 11884–11894.

Rothenberg, S.M., Ellisen, L.W., 2012. The molecular pathogenesis of head and neck squamous cell carcinoma. J. Clin. Invest. 122, 1951–1957. https://doi.org/10.1172/JCI59889

Sabatini, M.E., Chiocca, S., 2020. Human papillomavirus as a driver of head and neck cancers. Br J Cancer 122, 306–314. https://doi.org/10.1038/s41416-019-0602-7

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Res. 13, 2498–2504. https://doi.org/10.1101/gr.1239303

Shen, Y., Liu, J., Zhang, L., Dong, S., Zhang, J., Liu, Y., Zhou, H., Dong, W., 2019. Identification of Potential Biomarkers and Survival Analysis for Head and Neck Squamous Cell Carcinoma Using Bioinformatics Strategy: A Study Based on TCGA and GEO Datasets. BioMed Research International 2019, 1–14. https://doi.org/10.1155/2019/7376034

Singh, P., Gurung, R., Sultan, A., Dohare, R., 2023. Understanding the role of adipokines and adipogenesis family in hepatocellular carcinoma. Egypt J Med Hum Genet 24, 17. https://doi.org/10.1186/s43042-023-00401-5

Szklarczyk, D., Gable, A.L., Nastou, K.C., Lyon, D., Kirsch, R., Pyysalo, S., Doncheva, N.T., Legeay, M., Fang, T., Bork, P., Jensen, L.J., von Mering, C., 2021. The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Research 49, D605–D612. https://doi.org/10.1093/nar/gkaa1074

Tarazona, S., García, F., Ferrer, A., Dopazo, J., Conesa, A., 2012. NOIseq: a RNA-seq differential expression method robust for sequencing depth biases. EMBnet j. 17, 18. https://doi.org/10.14806/ej.17.B.265

Tutar, Y., 2014. miRNA and cancer; computational and experimental approaches. Curr Pharm Biotechnol 15, 429. https://doi.org/10.2174/1389201015051408281613335

Urashima, M., Hama, T., Suda, T., Suzuki, Y., Ikegami, M., Sakanashi, C., Akutsu, T., Amagaya, S., Horiuchi, K., Imai, Y., Mezawa, H., Noya, M., Nakashima, A., Mafune, A., Kato, T., Kojima, H., 2013. Distinct Effects of Alcohol Consumption and Smoking on Genetic Alterations in Head and Neck Carcinoma. PLoS ONE 8, e80828. https://doi.org/10.1371/journal.pone.0080828

Wu, Z., Yun-Tang, Cheng, Q., 2020. Data Mining Identifies Six Proteins that Can Act as Prognostic Markers for Head and Neck Squamous Cell Carcinoma. Cell Transplant 29, 096368972092930. https://doi.org/10.1177/0963689720929308

Yu, X., Wang, Z., Zeng, T., 2020. Essential gene expression pattern of head and neck squamous cell carcinoma revealed by tumor-specific expression rule based on single-cell RNA sequencing. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease 1866, 165791. https://doi.org/10.1016/j.bbadis.2020.165791

Zhang, B., Wang, H., Guo, Z., Zhang, X., 2019. A panel of Transcription factors identified by data mining can predict the prognosis of head and neck squamous cell carcinoma. Cancer Cell Int 19, 297. https://doi.org/10.1186/s12935-019-1024-6

Zoabi, Y., Shomron, N., 2021. Processing and Analysis of RNA-seq Data from Public Resources, in: Shomron, N. (Ed.), Deep Sequencing Data Analysis, Methods in Molecular Biology. Springer US, New York, NY, pp. 81–94. https://doi.org/10.1007/978-1-0716-1103-6_4