Oral Oncology Reports 7 (2023) 100084

Contents lists available at ScienceDirect

Oral Oncology Reports

journal homepage: www.journals.elsevier.com/oral-oncology-reports

Multi-omics approach reveals promising salivary protein markers for head and neck squamous cell carcinoma prognosis

Oriana Barros [a,b], Vito G. D'Agostino [c], Lucio Santos [b], Rita Ferreira [d], Rui Vitorino [a,b,e,*]

[a] Department of Medical Sciences, Institute of Biomedicine iBiMED, University of Aveiro, 3810-193, Aveiro, Portugal
[b] Experimental Pathology and Therapeutics Group, Research Center of IPO Porto (CI-IPOP) / RISE@CI-IPOP (Health Research Network) and Surgical Department of Portuguese Oncology Institute of Porto (IPO Porto) / Porto Comprehensive Cancer Center (Porto.CCC), Portugal
[c] Department of Cellular, Computational and Integrative Biology (CIBIO), University of Trento, Via Sommarive 9, 38123, Trento, Italy
[d] LAQV-REQUIMTE, Department of Chemistry, University of Aveiro, Campus Universitário de Santiago, 3810-193, Aveiro, Portugal
[e] UnIC, Department of Surgery and Physiology, Faculty of Medicine, University of Porto, 4200-319, Porto, Portugal

ARTICLE INFO

Keywords:
Head and neck squamous cell carcinoma
Transcriptomics
Salivaomics
Biomarkers
Prognosis

ABSTRACT

Head and neck squamous cell carcinoma (HNSCC) is one of the most common malignant cancers in the world. In advanced stages, the mortality rate is high. Stratification of HNSCC patients by molecular characteristics allows optimization of clinical management of these patients. Here, we explored the GEO and PRIDE Databases to identify putative prognostic biomarkers for HNSCC. Thus, an integrated transcriptome analysis was performed to identify the major differentially expressed genes in HNSCC tissues (using GSE12452, GSE13597, GSE31056, GSE6631 and GSE3524 GEO Datasets) in combination with a proteomic analysis to identify the overrepresented proteins in saliva samples from HNSCC patients (using PRIDE PXD012436 dataset). The panel of identified biomarkers was characterized using web tools, namely UALCAN, ToPP and PINAv3.0. From the combinatorial analysis of these Omics data, the salivary biomarkers with the greatest potential for clinical application for prognostic stratification were identified: ADH7, MMP9, and S100A14. Overall, this Multi-Omics approach provides comprehensive information on the potential value of combining ADH7, MMP9, and S100A14 as prognostic biomarker panel for risk stratification. Future studies should validate this panel of biomarkers in saliva samples from a large cohort of patients using targeted approaches, envisioning its translation into the clinical setting.

1. Introduction

Head and neck squamous cell carcinoma (HNSCC) is one of the most common malignant cancers. About 90% of Head and Neck Cancers are squamous cell carcinomas. The others are mostly adenocarcinomas, lymphomas, sarcomas, and melanomas. HNSCC usually affects men aged 40 and over, with marked smoking and alcohol habits and low socioeconomic status. This is an extremely mutilating type of cancer whose prognosis is poor when diagnosed in stages III/IV. About 60% are diagnosed in stage III-IVb with a 5-year overall survival (OS) of less than 50%, despite multimodal treatment (chemotherapy, radiotherapy, and surgery). When diagnosed in stage I/II, the disease is curable with radiotherapy or surgery, and the 5-year OS is 70–90% [1–5].

Multi-Omics profiling has been used to identify biomarkers in HNSCC and several sources of protein biomarkers have been studied, particularly body fluids (saliva, plasma, serum, and interstitial fluid) [6]. Saliva has gained attention as a novel liquid biopsy due to its proximity to the developing tumor. As a biological fluid, saliva has numerous advantages: non-invasive and pain free sample collection, repetitive sample collections, ease of sample processing at low cost. Saliva is a rich source of DNA, mRNA and proteins with potential biomarker value; however, its clinical utility in cancer set is limited due to some challenges. One of these challenges is the complexity of saliva molecular composition that comprises several non-tumorigenic molecules that mask the tumor-derived ones. For instance, the salivary proteome comprises more than 2000 proteins and peptides that are involved in the maintenance of oral cavity homeostasis, making difficult to identify tumor-derived proteins/peptides [7–9].

Several Omics-based biomarker signatures have been approved by the Food and Drug Administration (FDA). The first to be approved was the *MammaPrint*, developed at the Netherlands Cancer Institute to predict the risk of breast cancer recurrence based on DNA microarray gene

* Corresponding author. Department of Medical Sciences, Institute of Biomedicine iBiMED, University of Aveiro, 3810-193, Aveiro, Portugal.
E-mail address: rvitorino@ua.pt (R. Vitorino).

https://doi.org/10.1016/j.oor.2023.100084
Received 6 May 2023; Received in revised form 21 July 2023; Accepted 2 August 2023
Available online 10 August 2023

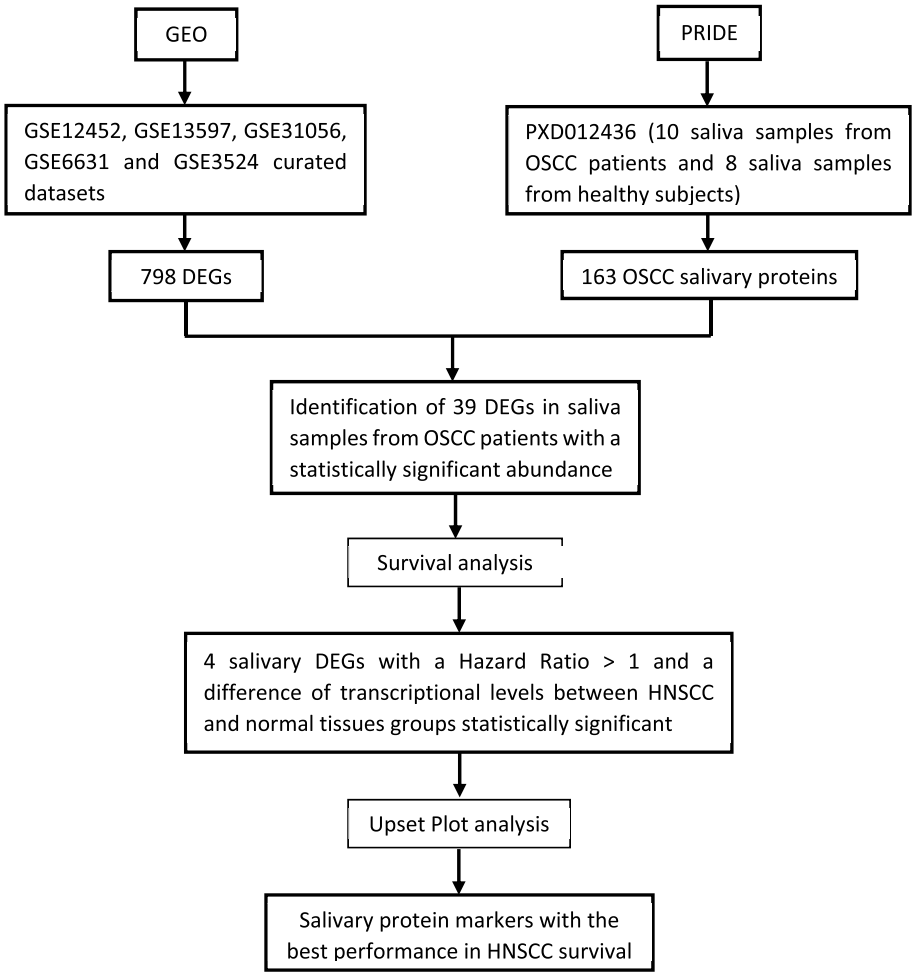

Fig. 1. Flowchart of the methodology followed in this study.

expression data. Other Omics-derived diagnostic or prognostic tests used in cancer that received FDA approval are the *Prosigna assay/PAM50* for predicting the risk of distant recurrence of breast cancer and the *PGDx Elio Tissue Complete* for evaluation of tumor mutation burden in solid tumors. However, these tests are performed using tissue samples that involve invasive techniques, highlighting the importance of developing strategies that allow the identification and validation of Omics-based biomarker signatures using biological fluids as source of liquid biopsy. Main challenges must be considered in study design: a sample size that ensures the statistical power of the study, rigor in the statistical analysis, well-defined objectives, normalization of individual Omics datasets, reproducibility, and interpretability of results. Despite the improvements on proteomics platforms, the identification of putative salivary protein markers of HNSCC remains a challenge. The combination of distinct Omics platforms might allow to overcome such methodological limitations. In the last few years with the use of mass spectrometry (MS)-based proteomics technologies, several repositories such as CPTAC, ICPC, PRIDE and PeptideAtlas have appeared. These databases can be used as discovery tools of novel protein targets by reanalyzing data from the datasets already known. The combination of the information extracted from proteomics approaches with other Omics provides a more integrated view of the molecular behavior of cells taking advantage of multiple molecular complexity levels increasing its potential for translation into clinical practice [10,11].

The goal of the present study was to identify putative salivary biomarkers through the combined analysis of data retrieved from transcriptomics analysis of HNSCC tissue samples with MS-based proteomics analysis of saliva samples from HNSCC patients, envisioning to improve the prognosis of this cancer type.

2. Methods

For the identification of the salivary protein markers with higher impact on the prognosis of HNSCC, a combinatorial Multi-Omics approach was applied in order to find the best panel of biomarkers. Firstly, transcriptomics data was revisited using GEO Datasets to identify the most overexpressed differentially expressed genes (DEGs) in HNSCC tissue. Then, proteomics datasets from PRIDE database were analyzed to find the proteins with the highest expression in saliva samples from HNSCC patients. Data from both analyses were intersected using a Venn Diagram to identify the most abundant proteins in the saliva of HNSCC patients encoded by the DEGs overexpressed in the tissue. From these biomarkers, those with the greatest impact on prognosis were selected and their interactome was studied to characterize the molecular and cellular mechanisms in which they are involved. The methodological approach used in the present study for the identification of salivary biomarkers with potential prognosis value for HNSCC is presented in Fig. 1.

Differentially expressed genes (DEGs) in HNSCC. Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/) database was used to obtain data from microarray platforms. GEO is a public repository at NCBI for storing high-throughput gene expression data. The gene expression profiles of the most promising salivary DEGs were extracted from GSE12452 (Affymetrix Hu133A GeneChips), GSE13597 (Affymetrix Hu133A GeneChips), GSE31056 (Affymetrix Hu133A GeneChips), GSE6631 (Affymetrix U95A GeneChips) and GSE3524

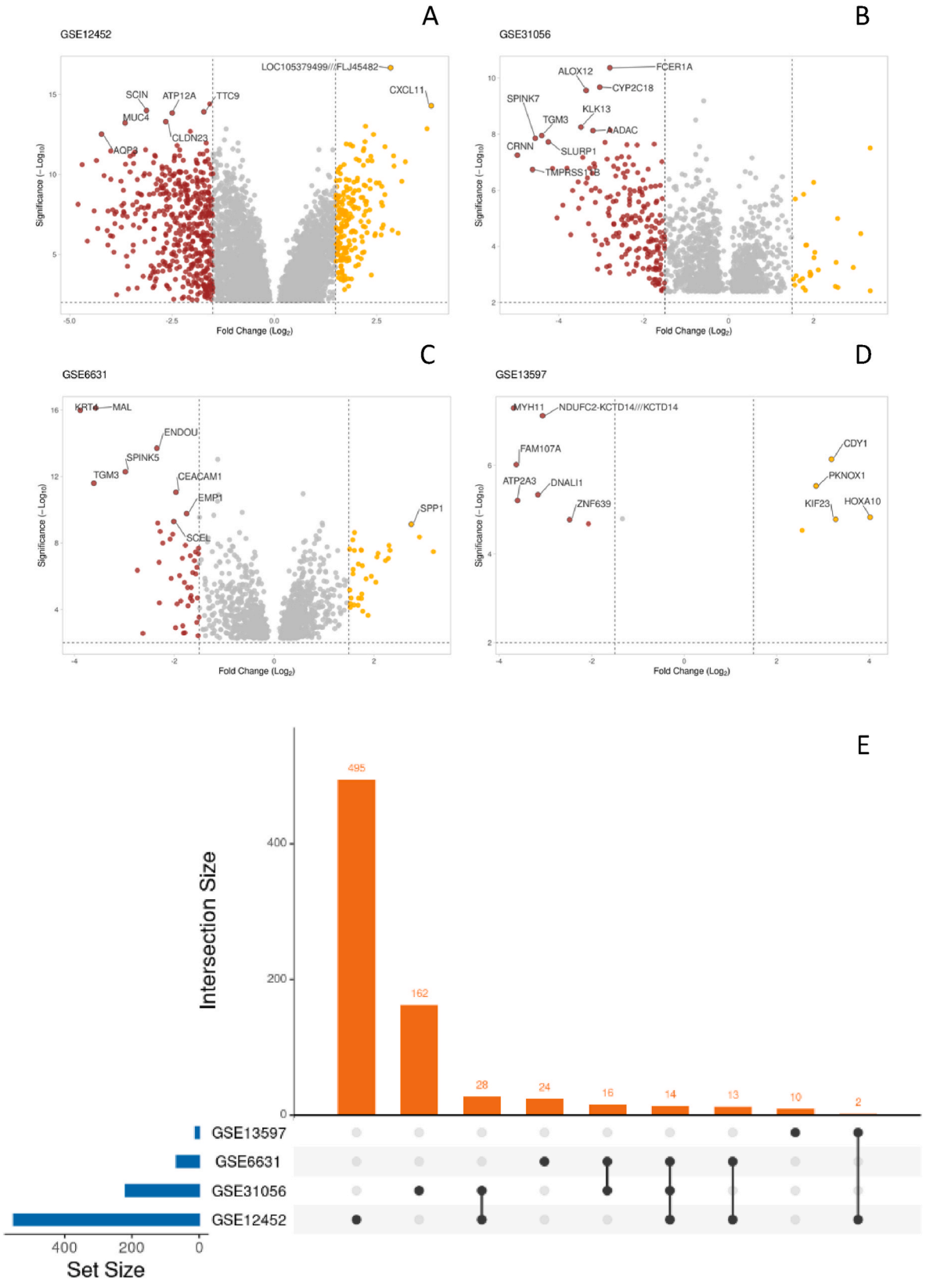

Fig. 2. Volcano plots of DEGs from GEO Datasets in VolcaNoseR. Volcano plots of the differentially expressed genes from the GSE12452 (A), GSE331056 (B), GSE6631 (C) and GSE13597 (D) datasets presenting the top 10 hits for each dataset. The annotated dots are the ten data points with the largest Manhattan distance from the origin. |Fold change (log2)| > 1.5 and significance > 2 were set as thresholds. Upset plot of DEGs overlapped in four datasets (E). In volcano plots, orange colour dots denote upregulated genes and brown colour dots denote downregulated genes. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 1
General information on GEO Datasets and platforms.

Dataset	HNSCC subtype	Number of Noncancerous Tissue Samples	Number of Cancer Tissue Samples	Upregulated genes	Downregulated genes	Platform
GSE12452	NPC	10	31	168	384	GPL570
GSE13597	NPC	3	25	5	7	GPL96
GSE31056	OSCC	24	72	26	194	GPL10526
GSE6631	HNSCC	22	22	24	43	GPL8300
GSE3524	OSCC	4	16	0	0	GPL96

Abbreviations: HNSCC, head and neck squamous cell carcinoma; NPC, nasopharyngeal carcinoma; OSCC, oral squamous cell carcinoma.

(Affymetrix Hu133A GeneChips) datasets from the GEO Database. These datasets compare the transcriptome identified in tissue samples from HNSCC patients with that of individuals without cancer. Only the GSE series shown in "Datasets" option were selected to ensure that only curated information was used. GEO2R (https://www.ncbi.nlm.nih.gov/geo/geo2r/) was performed to detect DEGs between HNSCC and normal tissue samples. A volcano plot was constructed for the DEGs obtained from each GEO Dataset using VolcaNoseR (https://huygens.science.uva.nl/VolcaNoseR/). The volcano plot shows the relationship between *P*-values and the magnitude of fold change in terms of control versus cancer. The criteria of DEGs were adjusted to significance ($-\log_{10}P$) > 2 and $|\log_2$ (fold change)| > 1.5. Co-expressed DEGs downregulated or upregulated were identified using an Upset plot (https://asntech.shinyapps.io/intervene/).

Salivary DEGs in HNSCC. To study which of the DEGs identified in the GEO Database could be present in saliva, the proteins of PXD012436 were extracted from the PRoteomics IDEntifications (PRIDE) Database. PRIDE is a repository of proteomic data obtained using MS techniques. The dataset PXD012436 contains the proteins extracted from saliva samples of oral squamous cell carcinoma (OSCC) patients and healthy subjects (considered as controls). Raw saliva samples prepared according to Lin et al. [12]. An equivalent amount of salivary proteins was separated and treated with 25 mM DTT and an equal volume of 10% SDS in 100 mM TEAB. The mixture was then incubated at 95 °C for 10 minutes. After cooling to room temperature, salivary proteins were alkylated in the presence of 50 mM iodoacetamide in the dark for 20 minutes. Protein identification and quantification: LC-MS /MS analysis was performed using an Ultimate 3000 nano- LC and a Q-Exactive mass spectrometer system (Thermo Scientific) according to a previously published protocol. The raw LCMS data were reprocessed for protein search (containing 82,678 protein sequences, release date: March 1,23). MS and MS /MS ion tolerances were set at 10 ppm and 0.02 Da, respectively. Only peptide and protein identifications with a false discovery rate (FDR) of 1% or less were included in the final data set. In cases where two or more proteins had the same set of identified peptides, only one major protein was reported.

MaxQuant software suite (version 2.0) was used for protein quantification, with most default settings adjusted. Label-free quantification was performed using the LFQ algorithm. Clustering and correlation analyses were performed in the Perseus environment (version 2.0.3.1) using embedded functions. Prior to analysis, MaxQuant generated label-free quantification (LFQ) intensities were log-transformed (base 2) and imputed with missing values according to the default settings in Perseus. Using a Venn diagram, the DEGs with potential to be detected in saliva samples of OSCC patients were identified. The expression profiles of these DEGs and the proteins encoded by the DEGs of interest were evaluated in UALCAN (TCGA and CPTAC). Furthermore, the impact of these salivary DEGs on the survival of HNSCC patients was also studied using ToPP to identify the best prognostic biomarkers. Prognosis-related genes were constructed using multivariate cox regression applying the following formula Prognostic Index (PI) = β. *n*, where β represents the cox coefficient and *n* represents the gene of interest. The performance of each model was assessed using Hazard Ratio (HR), *P*-value and CI 95%.

Interactome analysis for a deeper knowledge of the cellular and molecular mechanisms associated to the salivary DEGs. A PPI network was build using Protein Interaction Network Analysis (PINAv3.0) platform version 3.0 (https://omics.bjcancer.org/pina/queryProteinSet.action#) for integration of query proteins

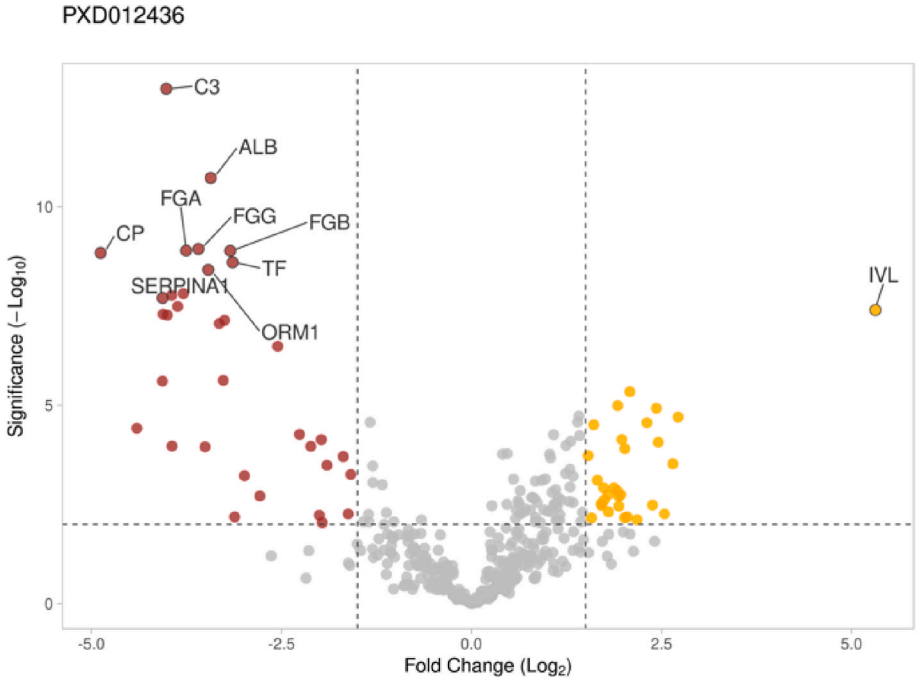

Fig. 3. Volcano plot of salivary proteins from PRIDE Dataset. Volcano plot of the expressed salivary proteins extracted from the PXD012436 dataset in VolcaNoseR. The ten proteins with the largest (Manhattan) distance from the origin above thresholds are highlighted. Orange dots are thus genes over-expressed in patients with oral cancer and brown dots are thus genes under-expressed. |Fold change (log2)| > 1.5 and significance > 2 were set as thresholds. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 2
Salivary DEGs common to the GEO and PRIDE Databases for HNSCC Datasets.

Genes	Proteins	GEO (log2)	PRIDE (-log *P*)	Tumour expression		Impact in HNSCC survival ($P < 0.05$)
				Gene	Protein	
A2ML1	Alpha-2-macroglobulin-like protein 1	−3.21	2.09	↓***	↓***	HR = 0.646 (0.463–0.903)
ACTA1	Actin, alpha skeletal muscle	−2.63	2.14	↓***	↓*	HR = 1.90 (1.290–2.790)
ADH7	Alcohol dehydrogenase class 4 μ/δ chain	−2.62	2.43	↓***	↓***	HR = 1.59 (1.13–2.25)
AMY1A	Alpha-amylase 1	−2.85	2.16	↓***	↓***	HR = 0.665 (0.477–0.926)
AMY2A	Alpha-amylase 2B			↓***	NA	ns
AMY2B	Pancreatic alpha-amylase			↓***	NA	HR = 0.576 (0.438–0.758)
ANXA1	Annexin A1	−2,01	2.46	↓***	↓***	ns
BPIFB1	BPI fold-containing family B member 1	−2.95	2.23	NA	↓***	NA
CALML3	Calmodulin-like protein 3	−1,78	4.57	↓**	↓***	HR = 0.658 (0.490–0.884)
CLCA4	Calcium-activated chloride channel regulator 4	−2.81	2.30	↓***	↓***	HR = 0.618 (0.474–0.807)
CNFN	Cornifelin	−2.16	2.16	↓***	NA	HR = 0.661 (0.482–0.908)
CRNN	Cornulin	−1.68	5.34	↓***	↓***	HR = 0.70 (0.518–0.946)
CSTA	Cystatin-A	−1.52	4.51	↓***	↓***	HR = 0.608 (0.464–0.796)
DSG3	Desmoglein-3	−2.14	3.11	↓**	ns	HR = 0.659 (0.463–0.939)
ECM1	Extracellular matrix protein 1	−1.55	4.07	↓***	↓***	HR = 0.736 (0.563–0.962)
GBP6	Guanylate-binding protein 6	−2.86	2.18	↓***	↓***	HR = 0.639 (0.489–0.836)
IL1RN	Interleukin-1 receptor antagonist protein	−2.46	2.80	↓***	ns	ns
IVL	Involucrin	−2.70	7.40	↓**	↓***	HR = 0.602 (0.397–0.914)
KLK11	Kallikrein-11	−2.17	3.05	↓***	↓***	HR = 0.7 (0.517–0.946)
KRT4	Keratin, type II cytoskeletal 4	−4.04	2.26	↓***	↓***	ns
KRT6A	Keratin, type II cytoskeletal 6A	−1.70	1.81	ns	↑***	ns
KRT6B	Keratin, type II cytoskeletal 6B	−1.70	2.61	ns	↑***	ns
KRT15	Keratin, type I cytoskeletal 15	−2.41	2.74	↓***	↓***	HR = 0.552 (0.321–0.949)
KRT78	Keratin, type II cytoskeletal 78	−3.67	1.81	↓***	↓***	HR = 0.7 (0.518–0.946)
LCN2	Neutrophil gelatinase-associated lipocalin	−3.67	1.71	↓***	ns	HR = 0.597 (0.359–0.994)
MMP9	Matrix metalloproteinase-9	1.76	2.04	↑***	↑***	HR = 1.42 (1.01–1.98)
MUC5AC	Mucin-5AC	−2.62	2.71	NA	↓***	NA
PKP1	Plakophilin-1	−1.71	2.16	ns	ns	ns
PPL	Periplakin	−2.10	4.70	↓***	↓***	HR = 0.567 (0.341–0.944)
RHCG	Ammonium transporter Rh type C	−1.71	4.73	↓***	↓***	HR = 0.686 (0.526–0.894)
S100A14	Protein S100-A14	−1.95	1.61	↓***	↓***	HR = 1.35 (1.01–1.80)
SAA1	Serum amyloid A-1 protein	−2.38	2.55	ns	↓***	HR = 1.45 (1.11–1.89)
SERPINB2	Plasminogen activator inhibitor 2	−3.33	3.72	↓***	↓***	HR = 0.597 (0.441–0.807)
SERPINB3	Serpin B3	−3.35	3.09	↓**	↓**	HR = 0.608 (0.370–0.998)
SLPI	Antileukoproteinase	−3.24	1.56	↓***	↓***	HR = 0.680 (0.501–0.924)
SPRR3	Small proline-rich protein 3	−3.01	2.71	↓***	↓***	HR = 0.622 (0.467–0.828)
TGM1	Protein-glutamine gamma-glutamyltransferase K	−3.04	1.65	↓***	↓***	ns
TGM3	Protein-glutamine gamma-glutamyltransferase E	−4.40	3.21	↓***	↓***	HR = 0.630 (0.426–0.931)

Abbreviations: *, *P*-value < 0.05; **, *P*-value < 0.01; ***, *P*-value < 0.001; HR: Hazard Ratio; NA: non available information; ns: not statistically significant.

and their interactor proteins from multiple databases into a single repository. This web tool has a function "Cancer Context" that allows the construction of a PPI network for a specified cancer type [13]. The following conditions were selected to perform this analysis: "TCGA-HNSC" dataset, tumor type specificity score > 2, Spearman correlation method, correlation coefficient 0.1 and survival comparison of High 50% and Low 50%. The tumor specificity score was calculated through comparison of the median expression level in each tumor to the median and interquartile range of its expression across all tumor types. After the identification of the interactor proteins related to our biomarkers of interest, we conducted a pathway enrichment using g:Profiler (https://biit.cs.ut.ee/gprofiler) to explore the function of the selected genes with biochemical, cellular, and molecular aspects. A *P*-value < 0.05 was defined as the threshold. g:Profiler is a toolset that relies on the Ensembl database as a main data source for enrichment analysis of the user's input gene list. This toolset identifies statistically significantly enriched biological processes, pathways, regulatory motifs, and protein complexes [14].

3. Results

For DEGs analysis, all datasets that fulfilled the criteria for inclusion in this study were download from GEO. For each dataset, genes with significance ($-\log_{10}P$) > 2 and |log fold change (FC)| > 1.5 were identified as significantly changed genes. All datasets displayed a set of DEGs, except GSE3524 dataset (no DEGs were found). We found 552 DEGs in GSE12452 (Supplementary Table S1), 12 in GSE13597 (Supplementary Table S2), 220 in GSE31056 (Supplementary Table S3) and 67 in GSE6631 (Supplementary Table S4). A volcano plot was constructed to exhibit the DEGs obtained from each dataset (Fig. 2A, 2B, 2C and 2D). Subsequently, an Upset plot analysis was performed to get the intersection of the DEGs (Fig. 2E) from all datasets. The overall information about each dataset is described in Table 1.

From the PRIDE repository, the PXD012436 dataset was used to extract the salivary proteins present in OSCC patients shown in Supplementary Table S5 and the corresponding Volcano Plot is shown in Fig. 3. The proteins encoded by the DEGs of interest found in statistically significant amounts in the saliva samples of OSCC patients are characterized in Table 2. Considering the results from Table 2, the DEGs with a HR > 1 and *P*-value < 0.05 (both in gene expression and survival analysis) were ACTA1, ADH7, MMP9 and S100A14, and were further studied to understand which is the best salivary proteins for prognosis of HNSCC. Thus, the gene expression profiles, and survival analyses are shown in Supplementary Fig. S1. Using CPTAC we also assessed the variation in protein expression of the proteins encoded by the DEGs of interest which are shown in Supplementary Fig. S2. ACTA1, ADH7 and S100A14 biomarkers showed to be under-expressed in HNSCC tissue and MMP9 was over-expressed in HNSCC tissue.

After intersecting the information extracted from the Upset Plot and the PXD012436 dataset (Table 3), ADH7 showed to be one of the biomarkers with greatest potential for prognostic monitoring among the various subtypes of HNSCC because it is the only biomarker from the panel of interest found in significant amounts in 3 of the 4 GEO Datasets used in this work. A gene enrichment analysis was performed for the best

Table 3
Overlapping DEGs in GSE12452, GSE31056, GSE6631 and PXD012436 datasets.

DEGs	Fold change (log_2)	Significance ($-log_{10}P$)	DEGs encoding proteins	DEGs encoding proteins function (according to Uniprot)
KRT4	−3.997	11.468	Keratin, type II cytoskeletal 4	Belongs to the intermediate filament family. Involved in cytoskeleton organization and epithelial cell differentiation
SERPINB2	−3.330	8.828	Plasminogen activator inhibitor 2	Belongs to serpin family, inhibits urokinase-type plasminogen activator
KRT13	−3.017	3.663	Keratin, type I cytoskeletal 13	Cytoskeleton organization
ADH7	−2.616	7.789	Alcohol dehydrogenase class 4 mu/sigma chain	Synthesis of retinoic acid by retinol oxidation
PPL	−2.097	6.081	Periplakin	May act as localization signal in PKB/AKT-mediated signalling
TMPRSS11D	−1.534	4.990	Transmembrane protease serine 11D	Host defense system on the mucous membrane

four salivary DEGs in prognosis of HNSCC and its interactors to provide a deeper understanding of the biological and molecular mechanisms associated with this set of DEGs. The results of this analysis are illustrated in Supplementary Fig. S3 and Supplementary Table S6. These DEGs are mostly involved in the formation and organization of the cytoskeleton and extracellular matrix. In PINAv3.0 the "Tumor-type specificity" analysis was conducted and is shown in Supplementary Fig. S4A. ADH7 was by far the protein with the highest median expression level in HNSCC when compared to the median and interquartile range of its expression across all tumor types included in this web tool. Subsequently, the "Pan-cancer View" analysis was performed in UALCAN for ADH7, MMP9 and S100A14 which is represented in Supplementary Figs. S4B–D. For the multivariate analysis, the biomarkers chosen were ADH7, MMP9 and S100A14 for their statistically significant impact on the prognosis of HNSCC and for the differential in terms of more marked expression in HNSCC tissue versus normal tissue. Using these three biomarkers, risk stratification was performed and a prognostic index was obtained for overall survival [PI_{OS} = (−0.0212 x ADH7) + (0.00264 x MMP9) + (−0.00431 x S100A14)], progression-free survival [PI_{PFS} = (−0.0354 x ADH7) + (−0.0726 x MMP9) + (−0.0426 x S100A14)], disease-specific survival [PI_{DSS} = (−0.0532 x ADH7) + (−0.0615 x MMP9) + (−0.0349 x S100A14)], disease-free interval [PI_{DFS} = (0.0599 x ADH7) + (0.0503 x MMP9) + (−0.0992 x S100A14)] and relapse-free survival [PI_{RFS} = (−0.0590 x ADH7) + (−0.1080 x MMP9) + (0.0102 x S100A14)]. Kaplan-Meyer plots corresponding to these parameters are shown in Fig. 4. In all survival analyses, this prognostic signature showed to be statistically significant for prediction of HNSCC prognosis.

4. Discussion

Based on information extracted from the GEO Database, 798 DEGs associated with HNSCC were identified. This information was cross-referenced with PRIDE data, which allowed the identification of 39 proteins encoded by the DEGs of interest present in saliva samples from OSCC patients. The role of these biomarkers in the prognosis of HNSCC was assessed using ToPP. After performing survival analysis, 4 salivary DEGs were selected: ACTA1, ADH7, MMP9 and S100A14 with a HR > 1 and *P*-value < 0.05. The transcript levels of the genes of interest were

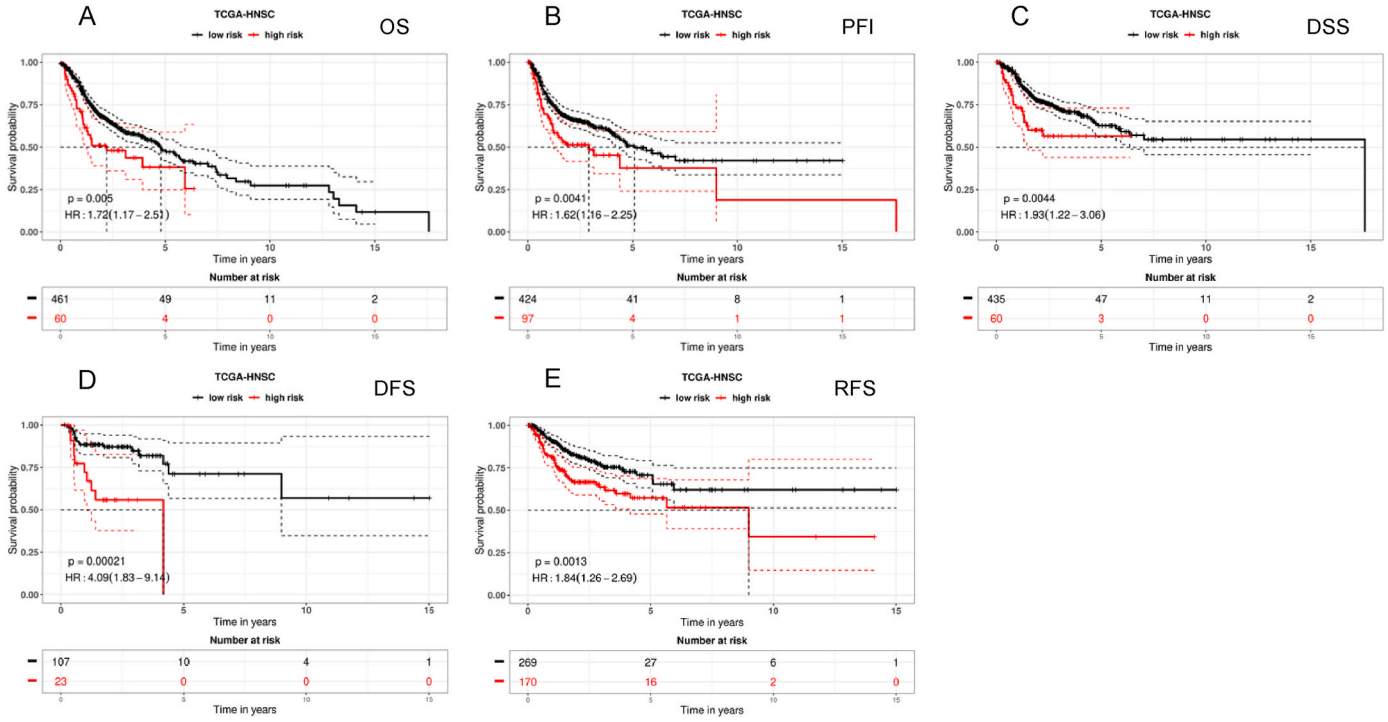

Fig. 4. Correlation between ADH7, MMP9 and S100A14 expression and survival analysis in ToPP. The Kaplan-Meier curves showed the statistical correlation between overall survival (A), progression-free interval (B), disease-specific survival (C), disease-free survival (D) and relapse-free survival (E) of the three-protein signature in HNSCC dataset. The results were all statistically significant (*P*-value < 0.05).

assessed using ToPP. The protein expression levels in healthy individuals and patients with HNSCC were analyzed using CPTAC. Biomarkers of interest were among the most overexpressed DEGs in the chosen GEO Datasets. From multivariate analyses made with the several possible combinations with ACTA1, ADH7, MMP9 and S100A14, the best profile for predicting survival in HNSCC patients was achieved for ADH7, MMP9 and S100A14. To better understand the role of ADH7, MMP9 and S100A14 in the cellular and molecular mechanisms of HNSCC, the interactome associated with the selected biomarkers was studied using PINAv3.0. With this analysis it was possible to observe that this set of biomarkers is strongly associated with the processes of cytoskeleton and extracellular matrix (ECM) formation and organization. ECM is responsible for supporting epithelial tissues and plays an important role in the regulation of gene transcription and cell signaling mechanisms. In HNSCC, the ECM can acquire an invasive phenotype responsible for the ability of tumor cells to interact with the ECM promoting the degradation of certain matrix components by proteolysis that subsequently migrate to the most susceptible ECM sites [15]. The regulatory processes of cytoskeleton formation and organization are associated with cancer cell motility and invasion. Both the ECM and the cytoskeleton are involved in the initiation, progression, invasion, and metastasizing processes associated with the development of HNSCC [16,17].

ADH7 is a member of the alcohol dehydrogenase family whose main function is the synthesis of retinoic acid, an essential hormone in the process of cell differentiation, being also involved in ethanol metabolism. Retinol is converted to retinaldehyde by ADH7 and then to retinoic acid by aldehyde dehydrogenase (ALDH), which binds to the retinoic acid receptor, playing an important role in the regulation of transcription of key genes during growth and development of organisms. The ADH-IV isoform has more affinity for retinol, which can be explained in part by the molecular weight of this molecule, since the active site of ADH7 is quite large, and the molecular weight of retinol is much higher than that of ethanol (molecular weight: 286.5 vs 46). However, when ethanol is present in abundant amounts it competitively inhibits the conversion of retinol to retinaldehyde, diminishing the synthesis of retinoic acid. Given that patients with HNSCC usually have marked ethyl habits, there is a reduced catalytic efficiency in retinol oxidation in this patient population. Regarding ADH7, Ren et al. studied a panel of metabolism related DEGs in HNSCC, and this gene was identified as one of the genes with the greatest impact on prognosis [18]. Li et al. studied the expression of some genes in patients with LSCC and confirmed that ADH7 is one of the most overexpressed genes in HNSCC [19]. According to the specificity analysis performed on PINAv3.0, ADH7 was shown to be the most specific protein for HNSCC. To date, there are no studies that have evaluated the levels and pathophysiological significance of ADH7 in fluids such as saliva, blood, or urine from patients with HNSCC.

The role of MMP9 as a predictive and prognostic biomarker in HNSCC is well documented. MMP9 is a metalloproteinase associated with the process of ECM degradation. Studies show that its overexpression is associated with a worse prognosis in patients with HNSCC. In patients who have been treated with radiotherapy, chemotherapy and/or surgery, high levels of MMP9 seem to be related to lower survival rate. A meta-analysis demonstrated that it is possible to monitor this biomarker using saliva as a source of liquid biopsy [20–23].

S100A14 encodes the S100 calcium binding protein A14 responsible for modulating p53/TP53 levels. Thus, it plays an important role in the regulation of cell survival and apoptosis. Pandey et al. demonstrated that loss of S100A14 expression is associated with reduced 10-year survival in OSCC patients. This protein seems to be involved in the regulation of the expression of some metalloproteinases such as MMP1 and MMP9 through a p53-dependent manner [24]. Chen et al. showed that there is a gradual reduction in S100A14 expression levels throughout the various stages of carcinogenesis. Since this protein controls cell proliferation by inducing G1-arrest, the loss of expression of this protein may be associated with the process of cell invasion [25].

The prognostic signature of the biomarker panel consisting of ADH7, MMP9, and S100A14 showed to be promising for HNSCC prognosis. This signature was statistically significant for survival analyzes OS, DSS, RFS, PFI, and DFS. Grønhøj et al. created a predictive prognostic nomogram for oropharyngeal squamous cell carcinoma that included p16 and other clinical and epidemiologic features [26]. Because the salivary biomarkers identified in this study have been shown to have a significant impact on the prognosis of HNSCC, optimization of the nomogram created by Grønhøj et al. by including these salivary proteins will allow the creation of a personalized risk score for each patient [26]. In addition, our salivary proteins have the potential to be translated into clinical practice by using saliva as a liquid biopsy for prognosis monitoring of HNSCC patients.There are several published works that have used GEO Datasets for biomarker identification, but no study to date has cross-referenced information extracted from GEO Datasets with information extracted from saliva samples from HNSCC patients [27–44]. In this work, we add insights on the identification of salivary biomarkers for HNSCC prognosis by selecting differentially expressed genes in tumors samples that are translated in proteins present in saliva at significant distinct levels. The main limitation of the study is that the PRIDE Dataset used to identify the most represented proteins in saliva samples from HNSCC patients has a small patient sample size. In PRIDE Database, there were other HNSCC Datasets, but this was the only one with clinical information available. Thus, to better support the biomarker value of the proposed panel of salivary proteins, future studies with larger patient cohorts are needed envisioning its translation into clinical practice.

5. Conclusion

With this study a novel protein-based signature for HNSCC prognosis using saliva as source of liquid biopsy was identified. This prognostic signature is composed by ADH7, MMP9 and S100A14 salivary protein biomarkers. ADH7 was validated *in silico* for the first time as a salivary prognostic biomarker of HNSCC, adding prognosis value to the proposed panel. For the translation of this protein-based signature for the clinical practice, studies with larger cohorts are necessary to validate this prognostic signature in the clinical setting.

Funding

This work was supported by the Portuguese Foundation for Science and Technology (FCT), European Union, QREN, FEDER, and COMPETE for funding Institute of Biomedicine (iBiMED) (UIDB/04501/2020, POCI-01-0145-FEDER-007628), LAQV-REQUIMTE (UIDB/50006/2020), Instituto Português de Oncologia do Porto Francisco Gentil, EPE (CI-IPOP-134-2020), COST Action INTERCEPTOR and by an individual scholarship FCT-UA-ECIU from Oriana Barros (University of Aveiro).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.oor.2023.100084.

References

[1] Johnson DE, Burtness B, Leemans CR, Lui VWY, Bauman JE, Grandis JR. Head and neck squamous cell carcinoma. Nat Rev Dis Prim 2020;6(1). https://doi.org/10.1038/s41572-020-00224-3.

[2] Canning M, Guo G, Yu M, et al. Heterogeneity of the head and neck squamous cell carcinoma immune landscape and its impact on immunotherapy. Front Cell Dev Biol 2019;7(APR). https://doi.org/10.3389/fcell.2019.00052.

[3] Rothenberg SM, Ellisen LW. The molecular pathogenesis of head and neck squamous cell carcinoma. J Clin Invest 2012;122(6):1951–7. https://doi.org/10.1172/JCI59889.

[4] Machiels JP, René Leemans C, Golusinski W, Grau C, Licitra L, Gregoire V. Squamous cell carcinoma of the oral cavity, larynx, oropharynx and hypopharynx: EHNS–ESMO–ESTRO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol 2020;31(11):1462–75. https://doi.org/10.1016/j.annonc.2020.07.011.

[5] Pfister DG, Spencer S, Adelstein D, et al. Head and neck cancers, version 2.2020. JNCCN Journal of the National Comprehensive Cancer Network 2020;18(7):873–98. https://doi.org/10.6004/jnccn.2020.0031.

[6] Wu Z hong, Cheng Q. Data mining identifies six proteins that can act as prognostic markers for head and neck squamous cell carcinoma, vol. 29; 2020. p. 1–8. https://doi.org/10.1177/0963689720929308.

[7] Punyadeera C, Slowey PD. Saliva as an emerging biofluid for clinical diagnosis and applications of MEMS/NEMS in salivary diagnostics. Nanobiomaterials in Clinical Dentistry 2019;(July):543–65. https://doi.org/10.1016/B978-0-12-815886-9.00022-X.

[8] Amenábar JM, da Silva BM, Punyadeera C. Salivary protein biomarkers for head and neck cancer. Expert Rev Mol Diagn 2020;20(3):305–13. https://doi.org/10.1080/14737159.2020.1722102.

[9] Hema Shree K, Ramani P, Sherlin H, et al. Saliva as a diagnostic tool in oral squamous cell carcinoma – a systematic review with meta analysis. Pathol Oncol Res 2019;25(2):447–53. https://doi.org/10.1007/s12253-019-00588-2.

[10] Misra BB, Langefeld C, Olivier M, Cox LA. Integrated omics: tools, advances and future approaches. J Mol Endocrinol 2019;62(1):R21–45. https://doi.org/10.1530/JME-18-0055.

[11] Creighton CJ. Clinical proteomics towards multiomics in cancer. Mass Spectrom Rev 2022. https://doi.org/10.1002/mas.21827. Published online.

[12] Lin YH, Eguez RV, Torralba MG, et al. Self-assembled STrap for global proteomics and salivary biomarker discovery. J Proteome Res 2019;18(4):1907–15. https://doi.org/10.1021/acs.jproteome.9b00037.

[13] Du Y, Cai M, Xing X, Ji J, Yang E, Wu J. PINA 3.0: mining cancer interactome. Nucleic Acids Res 2021;49(D1):D1351–7. https://doi.org/10.1093/nar/gkaa1075.

[14] Raudvere U, Kolberg L, Kuzmin I, et al. Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res 2019;47(W1):W191–8. https://doi.org/10.1093/nar/gkz369.

[15] Kumar KV, Hema KN. Extracellular matrix in invasion and metastasis of oral squamous cell carcinoma. J Oral Maxillofac Pathol 2019;23(1):10–6. https://doi.org/10.4103/jomfp.JOMFP_97_19.

[16] Kelley LC, Shahab S, Weed SA. Actin cytoskeletal mediators of motility and invasion amplified and overexpressed in head and neck cancer. Clin Exp Metastasis 2008;25(4):289–304. https://doi.org/10.1007/s10585-008-9154-6.

[17] Lee DY, Kang Y, Im NR, et al. Actin-associated gene expression is associated with early regional metastasis of tongue cancer. Laryngoscope 2021;131(4):813–9. https://doi.org/10.1002/lary.29025.

[18] Ren X, Song LJ, Zhang M, Zhang W, Lin K, He H. Identification of prognostic markers for head and NeckSquamous cell carcinoma based on glycolysis-related genes. Evid base Compl Alternative Med 2022:2022. https://doi.org/10.1155/2022/2762595.

[19] Li L, Wang R, He S, et al. The identification of induction chemo-sensitivity genes of laryngeal squamous cell carcinoma and their clinical utilization. Eur Arch Oto-Rhino-Laryngol 2018;275(11):2773–81. https://doi.org/10.1007/s00405-018-5134-x.

[20] Thangaraj SV, Shyamsundar V, Krishnamurthy A, et al. Molecular portrait of oral tongue squamous cell carcinoma shown by integrative meta-analysis of expression profiles with validations. PLoS One 2016;11(6). https://doi.org/10.1371/journal.pone.0156582.

[21] Ogbureke KUE, Weinberger PM, Looney SW, Li L, Fisher LW. Expressions of matrix metalloproteinase-9 (MMP-9), dentin sialophosphoprotein (DSPP), and osteopontin (OPN). At histologically negative surgical margins may predict recurrence of oral squamous cell carcinoma, vol. 3; 2012. www.impactjournals.com/oncotarget.

[22] Ruokolainen H, Pääkkö P, Turpeenniemi-Hujanen T. Serum matrix metalloproteinase-9 in head and neck squamous cell carcinoma is a prognostic marker. Int J Cancer 2005;116(3):422–7. https://doi.org/10.1002/ijc.21092.

[23] Shin YJ, Vu H, Lee JH, Kim HD. Diagnostic and prognostic ability of salivary MMP-9 for oral squamous cell carcinoma: a pre-/post-surgery case and matched control study. PLoS One 2021;16(3 March). https://doi.org/10.1371/journal.pone.0248167.

[24] Pandey S, Osman TA, Sharma S, et al. Loss of S100A14 expression at the tumor-invading front correlates with poor differentiation and worse prognosis in oral squamous cell carcinoma. Head Neck 2020;42(8):2088–98. https://doi.org/10.1002/hed.26140.

[25] Chen H, Yuan Y, Zhang C, et al. Involvement of S100A14 protein in cell invasion by affecting expression and function of matrix metalloproteinase (MMP)-2 via p53-dependent transcriptional regulation. J Biol Chem 2012;287(21):17109–19. https://doi.org/10.1074/jbc.M111.326975.

[26] Grønhøj C, Jensen DH, Dehlendorff C, et al. Development and external validation of nomograms in oropharyngeal cancer patients with known HPV-DNA status: a European Multicentre Study (OroGrams). Br J Cancer 2018;118(12):1672–81. https://doi.org/10.1038/s41416-018-0107-9.

[27] Chen Y, Li ZY, Zhou GQ, Sun Y. An immune-related gene prognostic index for head and neck squamous cell carcinoma. Clin Cancer Res 2021;27(1):330–41. https://doi.org/10.1158/1078-0432.CCR-20-2166.

[28] Shen Y, Liu J, Zhang L, et al. 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 Res Int 2019. https://doi.org/10.1155/2019/7376034. 2019.

[29] Ren Z, Zhang L, Ding W, et al. Development and validation of a novel survival model for head and neck squamous cell carcinoma based on autophagy-related genes. Genomics 2021;113(1):1166–75. https://doi.org/10.1016/j.ygeno.2020.11.017.

[30] Long Q, Huang C, Meng Q, et al. TNF patterns and tumor microenvironment characterization in head and neck squamous cell carcinoma. Front Immunol 2021;12. https://doi.org/10.3389/fimmu.2021.754818.

[31] Zhou C, Shen Y, Wei Z, et al. ITGA5 is an independent prognostic biomarker and potential therapeutic target for laryngeal squamous cell carcinoma. J Clin Lab Anal 2022;36(2). https://doi.org/10.1002/jcla.24228.

[32] Xu XL, Liu H, Zhang Y, et al. SPP1 and FN1 are significant gene biomarkers of tongue squamous cell carcinoma. Oncol Lett 2021;22(4). https://doi.org/10.3892/ol.2021.12974.

[33] Xie P, Wu S, Guo L, et al. Identification of candidate target genes and immune cells in oral squamous cell carcinoma. Comput Math Methods Med 2021:2021. https://doi.org/10.1155/2021/5802110.

[34] Yang W, Zhou W, Zhao X, et al. Prognostic biomarkers and therapeutic targets in oral squamous cell carcinoma: a study based on cross-database analysis. Hereditas 2021;158(1). https://doi.org/10.1186/s41065-021-00181-1.

[35] Song D, Tian J, Han X, Li X. A model of seven immune checkpoint-related genes predicting overall survival for head and neck squamous cell carcinoma. Eur Arch Oto-Rhino-Laryngol 2021;278(9):3467–77. https://doi.org/10.1007/s00405-020-06540-4.

[36] Jin Y, Qin X. Paired like homeodomain 1 and SAM and SH3 domain-containing 1 in the progression and prognosis of head and neck squamous cell carcinoma. Int J Biochem Cell Biol 2020;127. https://doi.org/10.1016/j.biocel.2020.105846.

[37] Zhang X, Yang M, Liu Y, et al. A novel 4-gene signature model simultaneously predicting malignant risk of oral potentially malignant disorders and oral squamous cell carcinoma prognosis. Arch Oral Biol 2021;129. https://doi.org/10.1016/j.archoralbio.2021.105203.

[38] Tian G, Fu Y, Zhang D, Li J, Zhang Z, Yang X. Identification of four key prognostic genes and three potential drugs in human papillomavirus negative head and neck squamous cell carcinoma. Cancer Cell Int 2021;21(1). https://doi.org/10.1186/s12935-021-01863-6.

[39] Wang L, Jia YP, Jiang ZY, Gao W, Wang BQ. FSCN1 is upregulated by SNAI2 and promotes epithelial to mesenchymal transition in head and neck squamous cell carcinoma. Cell Biol Int 2017;41(8):833–41. https://doi.org/10.1002/cbin.10786.

[40] Amiri-Dashatan N, Koushki M, Jalilian A, Ahmadi NA, Rezaei-Tavirani M. Integrated bioinformatics analysis of mRNAs and miRNAs identified potential biomarkers of oral squamous cell carcinoma. Asian Pac J Cancer Prev APJCP 2020;21(6):1841–8. https://doi.org/10.31557/APJCP.2020.21.6.1841.

[41] Wang J, Wang Y, Kong F, et al. Identification of a six-gene prognostic signature for oral squamous cell carcinoma. J Cell Physiol 2020;235(3):3056–68. https://doi.org/10.1002/jcp.29210.

[42] Shen T, Zhang J, Wang Y, Liu B. FCGBP is a promising prognostic biomarker and correlates with immunotherapy efficacy in oral squamous cell carcinoma. J Immunol Res 2022:2022. https://doi.org/10.1155/2022/8443392.

[43] Jin Y, Yang Y. Bioinformatics-based discovery of PYGM and TNNC2 as potential biomarkers of head and neck squamous cell carcinoma. Biosci Rep 2019;39(7). https://doi.org/10.1042/BSR20191612.

[44] Ding Z, Li H, Yu D. Development and validation of a hypoxia-related gene pair signature to predict overall survival in head and neck squamous cell carcinoma. Eur Arch Oto-Rhino-Laryngol 2021;278(10):3973–83. https://doi.org/10.1007/s00405-020-06580-w.