A novel 4 miRNA signature from salivary exosomes regulates multiple signalling pathways and predicts tumour progression in oral squamous cell carcinoma.

Shanaya Patel[a #], Aditi Patel[a #], Dushyant Mandlik[b], Kaustubh Patel[b], Krishna B.S Swamy[a], Vivek Tanavde[a,c*]

[a] Biological and Life Sciences, School of Arts and Sciences, Ahmedabad University, Ahmedabad, Gujarat, India

[b] Department of Head and Neck Oncology, HCG Cancer Centre, Ahmedabad, Gujarat, India.

[c] Bioinformatics Institute, Agency for Science Technology and Research (A*STAR), Singapore.

* **Corresponding author**

Vivek Tanavde PhD
Biological and Life Sciences,
School of Arts and Sciences,
Ahmedabad University,
Ahmedabad-380009
India
Tel: +91-91677088509
Email: vivek.tanavde@ahduni.edu.in

[#] **Both the authors contributed equally to the study**

Word count: 6596 words

Number of figures and tables: 9 figures and 1 table

Keywords: OSCC, oral cancers, salivary exosomes, exosomal miRNAs, cancer biomarkers, early detection

Abstract:

Overall survival rates of oral cancer patients have been low for over four decades. It is imperative to identify biomarkers that can be used for early diagnosis. In this study, we identified a miRNA signature from salivary exosomes of oral cancer patients for early detection of tumours. Additionally, we explored miRNA-mRNA regulatory networks and evaluated the functional role of these networks in oral carcinomas.

Four miRNAs (miR-30a, miR-140, miR-143, and miR-145) were differentially expressed in salivary exosome samples from OSCC patients compared to controls. These miRNAs showed significantly higher sensitivity, specificity and association with disease progression, recurrence, and chemo-resistance in OSCC patients. Network analysis of these genes conducted using Cytoscape, identified 16 hub proteins. Among these, *HIF1a*, *CDH1*, *CD44*, *EGFR*, and *CCND1* were the top hub genes known to drive oral cancer initiation and progression. Higher expression of these 5 genes showed a good correlation with poor survival in patients. We also identified biological processes and signalling cascades modulated by these miRNA-mRNA networks. Our findings reveal the immense diagnostic potential of these four miRNAs and uncover putative mechanisms by which dysregulation of these tumour suppressive miRNAs could effectively convert a normal epithelial phenotype into oral carcinoma.

Introduction:

Oral squamous cell carcinoma (OSCC) is one of the major causes of morbidity and mortality worldwide (1). For over four decades, OSCC has been associated with an extremely low five-year survival rate, mainly due to tumour detection at later stages. Conventional diagnostic approaches include invasive tissue biopsies that are challenging, expensive, painful, difficult when accessibility to deeper sites is required, and unable to predict tumour heterogeneity (2, 3). Despite recent advances in molecular diagnostics, no disease-specific biomarkers are clinically available for early risk prediction of OSCC. Therefore, it is important to identify robust biomarkers that are detectable using non-invasive liquid biopsy techniques to facilitate the early diagnosis of oral cancer.

Tumour-derived exosomes (40-150 nm in diameter) are secreted in various body fluids like blood, saliva, urine, cerebrospinal fluid, amniotic fluid and are useful sources of liquid biopsies. These exosomes mediate intercellular communication through transmission of active biomolecules (mRNA, microRNA (miRNA), DNA, proteins, and lipids) into the tumour microenvironment. These exosomes are an attractive target for early diagnosis and risk prediction of cancer as they represent the molecular and mutational landscape of the tumour bulk (4). Given that saliva has close proximity to the oral cavity and can be collected using convenient non-invasive and cost-effective methods, it is a good source of exosomes produced in oral cancers compared to other body fluids (5).

Recently, interest in the identification and characterization of exosomal RNAs, particularly miRNAs, has increased, as they are highly tissue-specific, can be effectively transported to donor cells, and negatively regulate 60% of protein-coding genes in humans (6). Moreover, exosomal miRNAs have demonstrated immense clinical potential in early risk prediction, assessment of disease progression, real-time monitoring of therapies, and identification of therapeutic targets in various malignancies (7). Hence, it is necessary to identify miRNAs for early diagnosis of oral cancer and elucidate the miRNA-mRNA networks leading to deregulation of various biological mechanisms and signalling pathways (8).

In this study, we conducted a comprehensive comparative analysis of miRNA and mRNA profiles of tumour tissue and salivary exosomes derived from leukoplakia (pre-malignant) and oral cancer patients and normal controls. We identified miRNA-mRNA gene regulatory networks that play an

important role in the early diagnosis of oral carcinoma. The results provide insights into molecular mechanisms responsible for transforming a normal epithelial phenotype into a malignant one. To the best of our knowledge, this is the first study to explore the correlation between saliva and tumour derived miRNA-mRNA regulatory networks, which can eventually facilitate early diagnosis of oral cancer.

Materials and Methods

Patient group.

Resected tumour tissue and unstimulated saliva samples were collected from patients diagnosed with OSCC (excluding tongue, larynx, pharynx, and hypopharynx) at the HCG Cancer Centre, Ahmedabad. The patients were authorized for the primary surgical resection by the institutional team and had malignancies confirmed by the pathologist were included in the study. HPV, HIV, HCV and HbsAg positive malignancies, paediatric cases and patients initially treated with other adjuvant therapeutic modalities were excluded from the study. All patients provided written informed consent for their participation in the study. This study was reviewed and approved by the HCG Cancer Centre's Ethics Committee (ECR/92/Inst/GJ/2013/RR-16) for human subject research and the use of human tissues. This study also complied with the guidelines set forth by the Declaration of Helsinki (2008).

Tissue and saliva collection:

Samples were collected only during R0 resections of early and late stage OSCC; the samples for biopsy were taken from the tumour tissues. Resected specimens were processed for histopathological evaluation and residual tumour samples were snap-frozen and stored at -80 °C for further processing. For normal controls, brush biopsy samples were taken from healthy individuals (matched for age and sex with patients) with etiological history of smokeless tobacco and no clinically detectable oral lesions.

Whole saliva was collected in sterile tubes from healthy individuals and OSCC patients prior to surgical resection, using a protocol described in a previous study (9). To ensure that unstimulated saliva samples were collected, the patients were instructed not to forcefully expectorate/discharge into the collection tubes. The saliva samples were kept at -80 °C until further use. A total of 23

patient derived samples and 21 samples from age- and sex-matched healthy volunteers were sequenced. The details of sequencing samples are given in Table 1. Buccal scrapings and salivary exosomes from multiple healthy volunteers were pooled (3 volunteers/pool) to generate reference sequences. It was difficult to harvest enough miRNA and mRNA for sequencing from the buccal scrapings of healthy controls. This was not an issue in samples derived from patient derived tumour tissue or salivary exosomes. Moreover, pooling of healthy volunteers gets rid of individual variations in the miRNA and mRNA transcripts from these volunteers. Thus, these pooled sequences can be treated as representative reference sequences of healthy volunteers. Sequencing results were validated by qRT-PCR using a total of 80 samples comprising 35 tumour tissue samples, 15 leukoplakia tissue samples, 5 control buccal scrapings (3 volunteers/pool), 20 salivary exosome samples from OSCC patients, and 5 salivary exosome samples from healthy volunteers.

Table 1: Patient sample and analysis details

Sample type	**Sampling details**	**Methodology used**
Testing Cohort (n=30)		
Tumour and Leukoplakia Tissue	15	Small RNA and mRNA Sequencing
Buccal Scrapings	4 pooled samples (3 volunteers/pool)	
Saliva	8	Small RNA Sequencing
Saliva from healthy individuals	3 pooled samples (3 volunteers/pool)	
Validation Cohort (n=80)		

Tumour and Leukoplakia Tissue	50	Real time PCR
Buccal Scrapings	5 pooled samples (3 volunteers/pool)	
Saliva	20	
Saliva from healthy individuals	5	

Exosome isolation:

Saliva from OSCC patients and healthy controls was centrifuged at 2000 × *g* for 10 min at 37°C to remove cells and debris. The supernatant was transferred to a sterile tube, and the exosomes were precipitated using a commercially available kit as per the manufacturer's protocol (Invitrogen™ Total Exosome Isolation Reagent (from other body fluids), Thermo Fisher Scientific, Waltham, MA, USA). The exosome pellet was either re-suspended in 50µL of phosphate-buffered saline (PBS) for characterisation or in 750µL of TRIzol LS reagent (Thermo Fisher Scientific) for storage at - 80 °C for subsequent RNA extraction.

Exosome characterization:

The exosome pellet was diluted in physiologic solution (1:500) and analysed by NanoSight LM10 (Malvern Instruments Ltd., Malvern, UK). The concentration (particles/mL) and size of exosomes were measured using the Nanoparticle Tracking Analysis (NTA) software (Malvern Instruments Ltd). Negatively stained exosomes were subjected to transmission electron microscopy (TEM) to determine particle size. The exosome specimens were placed onto freshly glow-discharged 300 mesh carbon formvar grids and set for 10 min. The solution was absorbed with filter paper,

following which 5µL of 2% glutaraldehyde was added to the grids and retained for 3 min. The grids were then wick-dried, rinsed with water, and stained with 2% uranyl acetate for 1 min. Each grid was then individually air-dried for 10 min before examination using the JEOL JEM1400 Plus transmission electron microscope (Jeol, Tokyo, Japan) at 100 kV. Images were obtained via a chamberscope film camera.

Expression of tetraspanins on isolated exosomes was investigated by flow cytometry. The isolated exosomes were stained with anti-human CD9 allophycocyanin (CD9-APC), anti-human CD63 Alexa Fluor 488 (CD63-Alexa 488), and anti-human CD81 phycoerythrin (CD81-PE antibodies (Thermo Fisher Scientific, Waltham, MA, USA). The exosomes were diluted in PBS and incubated for 30 min at room temperature antibody solutions having predetermined optimal concentrations. The samples were then analysed using the FACS Calibur™ Flow Cytometer (Becton Dickinson Biosciences, San Jose, CA). Particle coincidence was evaluated by acquiring different dilutions of stained exosomes. Events/second were recorded at each dilution using the high sample flow rate on the instrument. The lowest dilution with which a linear correlation could be observed between dilution and events/second was considered optimal.

RNA extraction:

We have modified the method described by Prendergast et al. for RNA extraction from exosomes (10). RNA was extracted by adding 750µL TRIzol LS reagent (Thermo Fisher Scientific) and 200µL chloroform to each salivary exosome sample (40µL); the samples were mixed thoroughly by vortexing for over 30 s and incubated at room temperature for 10 min. Phase separation was performed by centrifugation of the samples at 12,000 × g at 4 °C for 15 min. The upper aqueous phase was collected, and a volume of sodium acetate (3 M, pH 5.5) equal to 10% of the sample volume was added to each sample, following which 5µL glycogen (5 mg/mL, Sigma Aldrich, St. Louis, MO) and 2.5 times the acquired aqueous volume of absolute ethanol were added. The samples were mixed, incubated overnight at -80 °C, and then centrifuged at 16,000 × g at 4 °C for 30 min to pellet the RNA. The pellet was washed with 70% ethanol and centrifuged at 16,000 × g at 4 °C for 5 min. Then ethanol was aspirated, and the pellet was air dried for 5 min before being resuspended in 32µL of nuclease-free water. Purified RNA was quantified using Qubit 4 Fluorometer (Thermo Fisher Scientific, Waltham, MA, United States).

The total RNA was extracted from the tissue samples of OSCC patients using TRIZOL reagent

following the manufacturer's instructions. The RNA yield and purity were determined using Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, United States) and Qubit 4 Fluorometer (Thermo Fisher Scientific, Waltham, MA, United States).

miRNA sequencing:

Genome-wide miRNA expression profiles of tumour tissues and salivary exosomes derived from OSCC and patients were generated by miRNA-sequencing at Genotypic Technology Pvt. Ltd., Bengaluru, India. Small RNA sequencing (smRNA) libraries were prepared using QIAseq® miRNA library kit protocol (QIAGEN, Maryland, U.S.A.) as per manufacturer's instructions. Total RNA (5µL) was used as the starting material. Specific 3′-OH groups of miRNAs were ligated with 3′ adapters, and subsequently with 5′ adapters. The adapter ligated fragments were reverse transcribed using Unique Molecular Index assignment by priming with reverse transcription primers, and the complementary DNA (cDNA) thus formed was enriched and barcoded by PCR amplification. The Illumina-compatible sequencing libraries were quantified using Qubit 4 Fluorometer (Thermo Fisher Scientific, MA, USA). The resultant miRNA libraries were sequenced on the Illumina NextSeq 500 platform (Illumina, CA, USA), generating around 12-15 million single end reads per run for each sample.

The overall quality of raw reads was evaluated using FastQC program (11). Raw reads were trimmed to remove adapter sequences using Cutadapt script (v. 1.8) with the following options: --trim-n -a file: 'sample specific adapter sequences in FASTA file' -m 15 -q 20, where "-m" denotes the minimum fragment length in bp and "-q" denotes the minimum read quality score encoded in ASCII (phred quality + 33) format (12). Sequencing reads that passed quality control were mapped to the human reference genome (hg38) available from NCBI using the miRDeep2 (v0.0.7) package (13). The processed reads from all FASTQ files were mapped by the mapper.pl script of the miRDeep2 tool. This procedure generated non-redundant reads with the counts, source, and genomic location of the mapped reads. The miRDeep2 tool was then used to annotate putative novel miRNAs as well as validate predicted miRNAs from the miRBase database using mapped reads, and miRNAs predicted by human miRBase. The miRNAs were considered expressed when detected with an miRDeep2 score ≥ 5, quality score >30 reads/sample, RPM normalized expression value ≥ the first quantile of expression distribution, false discovery rate (FDR) of ≤ 0.05, log2 fold change expression of > 2, and statistically significant abundance in each sample

(adj p-value < 0.05). This script produced a raw count file that was used as input for the DESeq2 R/Bioconductor package for further analysis (Figure 1).

Transcriptome sequencing:

Transcriptome profiling of tumour tissues and salivary exosomes derived from OSCC patients and controls was performed by RNA-seq at Genotypic Technology Pvt. Ltd., Bengaluru, India. Library preparation for transcriptome sequencing was performed using NEBNext® Ultra™ II directional RNA library prep kit and sequencing was performed on Illumina HiSeq 2500 platform using 150 × 2 chemistry to generate 20 million reads per sample per run. Twelve isolated total RNA samples (~2.5 µg) were used to generate cDNA libraries using Illumina TruSeq RNA sample preparation v2 kit. Paired end sequencing (2 × 150 bp) was performed on the cDNA libraries. RNA-Seq generated over 40 million reads with Q30 > 90% per sample on average. Raw RNA-seq reads were first inspected for overall quality by FastQC and then cleaned by Cutadapt script (11, 12). The cleaned reads were mapped to human genome hg38 using the STAR aligner (v2.4.0h) (14). The mapped reads were computed using StringTie (v1.3.0) to improve transcript assembly (15). After transcript assembly for each sample, the resulting transcriptomes were combined with 'stringtie --merge' into a single representative transcriptome, which was followed by the estimation of the expression level. Finally, gene-level quantification was performed using the feature Counts which is a reads summarization program that can efficiently count the mapped reads for genomic features (i.e., genes and exons) (16) (Figure 1).

Differential expression analysis:

Differential expression tests in DESeq2 were conducted for the samples possessing miRNA genes with read counts ≥30. The DESeq function in DESeq2 was used to estimate size factors and dispersions and fit a model. The DeSeq2 likelihood ratio test (LRT) was used to compare the conditions (leukoplakia - control) - (tumour - control) and obtain differentially expressed miRNAs. The results were visualized using the DEGreport tool in the Bioconductor package. The criteria for miRNAs to be considered differentially expressed were as follows: log2 fold-change > 2 and the adjusted p-value < 0.01.

miRNA-specific quantitative reverse transcriptase PCR

Expression of the candidate miRNAs selected by means of differential expression analysis (log2 fold change > 2; adj p-value < 0.01) was validated by SYBR Green PCR assay. U6 snRNA was used as an endogenous control. Total RNA was reverse transcribed using the miScript II RT kit (QIAGEN, Frederick, MD, USA) according to the manufacturer's instructions. The reaction mixtures were subjected to incubated for 60 min at 37 °C, followed by incubation for 5 min at 95 °C to inactivate reverse transcriptase, and then stored at -80 °C until further use. Quantitative RT-PCR was performed using the miScript SYBR® Green PCR kit (QIAGEN, Frederick, MD, USA) on the QuantStudio 5 Real-Time PCR system (Applied Biosystems, Foster City, CA, USA). Relative expression of individual miRNAs in each case was calculated using the $2^{-\Delta\Delta Ct}$ method. All PCR reactions were carried out in triplicates.

Statistical analyses were performed using the SPSS (v19.0) software (SPSS, Inc. Chicago, IL) and the GraphPad prism 5.0 software (17). The results of the student's *t*-test are presented as mean ± standard deviation (SD). All experiments were performed in triplicates. Statistical significance was set at p-value < 0.05.

The discrimination power of the differentially expressed miRNAs as biomarkers for early OSCC diagnosis was evaluated by the Receiver Operating Characteristic (ROC) curves. The ROC curves were constructed with the relative miRNA expression levels of controls and OSCC patients using the demo version of GraphPad Prism 6.01 software, which calculated the sensitivity, specificity, area under curve (AUC), and p-value. The optimal threshold value was decided using Youden's index (sensitivity + specificity -1).

Different miRNA subsets were subjected to principal component analysis (PCA) for investigating relationships between samples, and the silhouette score was used to estimate the optimal number of clusters. We used the cluster package in R to calculate the silhouette score with Pearson correlation coefficients as distance measures (18).

miRNA target prediction

Candidate miRNA-mRNA target relationships were predicted by at least one or more of the following target prediction algorithms (union set) extracted from miRDB (19), TargetScan 5.1 (20), and miRWalk (21). Genes were considered differentially expressed if log2 fold change > 2, and adj p-value was ≤ 0.05.

Interaction networks

Based on the differentially expressed genes, protein-protein interaction networks were generated using the STRING database (22). Network visualization was performed using the open-source software platform Cytoscape (v3.6.1) (23). Further, these differentially expressed genes were also imported into Ingenuity Pathways Analysis software (IPA) (QIAGEN Inc., https://digitalinsights.qiagen.com/IPA) and were subjected to regulatory network analysis (24).

Pathway analysis

We also used IPA (QIAGEN) to identify significant pathways based on differentially expressed genes. The list of top differentially expressed genes containing gene identifiers and corresponding expression values, was uploaded into the IPA software. For canonical pathway analysis, the -log(p) >2 was taken as the threshold, and Z-score >2 was defined as the threshold of significant activation, whereas Z-score <-2 was defined as the threshold of significant inhibition.

Results

Identification and characterisation of salivary exosomes derived from OSCC patients

Salivary exosomes were isolated from patients diagnosed with leukoplakia and OSCC and healthy controls. These exosomes were characterised by NTA, TEM, and flow cytometry. The NTA results showed a single peak for the concentration of exosomes in the size range of 30 - 40 nm (Figure 2 a, b). Salivary exosomes derived from OSCC patients had a mean size of 31.5 ± 1.9 nm and a concentration of 7.31×10^8 particles/mL. The control exosomes had an average particle size of 39.8 ± 4.2 nm and a concentration of 6.67×10^7 particles/mL. The exosomes isolated from OSCC patients were smaller in size and their saliva contained 10-fold more exosomes than that of the controls. TEM results (Figure 2 c, d) represent sphere-shaped vesicles with a mean radius of 40-50 nm, which was consistent with the NTA profiles. Further, flow cytometry analysis indicated the presence of the tetraspanins CD9, CD63, and CD81 (86-88% exosomes expressed these markers) in salivary exosomes from both patients and healthy counterparts (Figure 2 e–i), thus confirming that the isolated vesicles comprise pure exosomal population.

Differentially expressed miRNAs enriched in both salivary exosomes and primary tumours from OSCC patients identified by miRNA profiling

RNA-Seq analysis of small RNAs was performed to identify differentially expressed miRNAs in tumours and salivary exosomes derived from OSCC patients compared to those derived from age- and sex-matched controls. For tumour tissue samples from OSCC patients, the mean RNA integrity value was 7.6 with a 260/280 ratio ranging from 1.87–2.12 (average 2.06) and an average RNA concentration of 655.02 ng/µL across all samples. However, salivary exosomes had a mean RNA integrity value of 5.6 (260/280 ratio: 1.68-2.01) and an average concentration of 57.31 ng/µL. The low integrity value of RNA extracted from salivary exosomes could be due to the RNA-degrading enzymes present in saliva.

An average of 25.4 million reads per sample were obtained for all tissue and salivary exosomes samples. After pre-processing, adapter sequence trimming and filtering low quality reads, the mean Q20 values for tissue-based libraries and salivary exosome-derived libraries were 97.6% (2,31,53,568) and 93.40%, respectively. Reads shorter than 15 nucleotides in length and a quality score <30 (reads less than 30) were excluded to ensure high quality of the sequencing results. Amongst these, an average of 68% (range 61-75%) of the reads could be mapped to the reference human genome (hg38) with high confidence.

188 miRNAs were differentially expressed (log2 fold change $\pm$ 2; adj p-value < 0.05) in tumour tissues, 181 miRNAs in leukoplakia samples, 316 miRNAs in salivary tumour exosomes and 69 miRNAs in salivary leukoplakia exosomal samples compared to those in their normal counterparts. Amongst these, 47 miRNAs were found to be differentially expressed across all sample types (saliva and tissue) and stages, including pre-cancerous lesions (Figure 3a). A comparison of our results with the data from The Cancer Genome Atlas (TCGA) database showed that 27 of these differentially expressed miRNAs exhibited significantly differential expression between both datasets (Figure 3b). Based on a stringent cut-off for adj p-value (p<0.01) and log2 fold change > 2), 12 miRNAs (10 downregulated and 2 upregulated) were found to be robustly differentially expressed across all sample types. These 12 miRNAs were further validated by means of PCR with a validation set comprising a larger patient cohort (Table 1). Seven miRNAs (miR-140, miR-143, miR-145, miR-30a, miR-21, let-7i, miR-423) were found to have significant differential

expression in both tissue and salivary exosome samples derived from OSCC patients compared to those derived from controls (Figure 3c). This indicates that differentially expressed miRNAs enriched in salivary exosomal fractions could reflect the miRNA profile of the primary tumour site.

An early diagnostic predictive model based on miRNA expression

The seven significantly expressed miRNAs were subjected to ROC analyses both individually and collectively to determine their utility as potential early diagnostic markers and generate a predictive model based on miRNA expression and verify its sensitivity and specificity. On assessing various combinations of miRNAs, miR-30a, miR-140, miR-143, and miR-145 showed an AUC of 0.92 ($p < 0.0001$) with 92% sensitivity and 98% specificity in tissue samples as compared to the seven miRNAs in the validation set of tissue samples from both OSCC and leukoplakia patients, which had an AUC of 0.8 ($p < 0.0001$) with 80% sensitivity and 85% specificity (Figure 4a). Similarly, ROC analysis of salivary exosomes revealed an AUC of 0.97 ($p < 0.0001$) with 95% sensitivity and 98% specificity for the 4-miRNA combination model, which was higher than that of the 7-miRNA model and in accordance with the findings for the tissue samples (Figure 4b). These results indicate the potential utility of this 4-miRNA model having similar efficacy irrespective of the sample type (saliva or tissue) in effectively differentiating the cancerous and precancerous patients from the control counterparts. Further, PCA models were constructed for both salivary exosomes as well as tissue derived from oral cancer patients compared to their healthy counterparts using miR140, miR143, miR145 and miR30a expression levels. These PCA plots demonstrated that leukoplakia and tumour miRNA profiles in both exosomes and tissues were clustered differently compared to those in control samples. Interestingly, among all profiled miRNAs, patient-derived miRNAs clustered separately. Thus, exosome-derived miRNAs can be used to differentiate between non-cancerous and cancerous cohorts (Figure 4c).

Four-miRNA signature correlates with disease progression, therapeutic refractoriness, and relapse

Expression levels of the four significantly expressed miRNAs (miR-140, miR-143, miR-145, and miR-30a) were correlated with OSCC tumour stages using a student's t-test to assess the feasibility and effectiveness of these miRNAs as biomarkers. Characteristic stage-dependent

variation in the expression levels of tissue miRNAs and salivary exosomal miRNAs was analysed between various stages (I, II, III, and IV) of OSCC. Significant upregulation was observed in the expression of miR-30a, miR-140, and miR-145 in tissue samples (Figure 5a) and of miR-30a, miR-140, and miR-143 in salivary exosomal samples (Figure 5b) in early stages (stage-I and II; n = 20) compared to that in late-stage patients (stage-III and IV; n = 30) in tissue as well as saliva samples ($p < 0.05$).

We further examined whether differential expression of the identified miRNAs had any clinical association with the relapse and chemo-resistance in patients using Student's t-test. We found that miR-140, miR-145 and miR-143 showed significantly downregulated expression ($p<0.05$) in recurrent OSCC tumour tissue (n = 8) samples as compared to that in non-recurrent samples (n = 8) (Figure 5c). Similarly, all the four miRNAs included in our predictive model had significantly downregulated expression in the chemo-resistant (n = 5) cohort compared to that in the chemo-sensitive patient cohort (n = 5; $p < 0.05$) (Figure 5d). Among these, the expression of miR-140, miR-143, and miR-145 can potentially predict disease progression, recurrence, and therapeutic effectiveness in OSCC. However, further validation of this hypothesis by means of multi-centric trials involving larger patient cohorts is warranted to test this hypothesis.

RNA-seq analysis-based identification of differentially expressed genes in tissues from OSCC patients compared to those in tissues from normal controls

Matched samples of oral cancer patients were subjected to small RNA sequencing using high-throughput RNA sequencing to understand the transcriptome level mechanism and elucidate the functional relevance of the identified miRNAs in regulating oral cancer progression. After pre-processing, adapter sequence trimming, and filtering low quality reads, 17.8–28.5 million reads (an average of 94.6% clean reads) were obtained from various tumour samples. Among these, an average of 62-67% of the reads mapped to the reference human genome (hg38) with high confidence. Transcriptome analysis identified 10660 and 11632 differentially expressed mRNAs in tumour and leukoplakia samples of OSCC patients, respectively, in comparison to those expressed in their non-cancerous counterparts. On comparing these results with the TCGA datasets, we found a total of 5868 mRNAs that were significantly expressed (log2 fold change > 2; adj p-value < 0.01) across these datasets irrespective of the varying clinicopathological status of patients and differences in sequencing techniques.

Prediction of miRNA targets and identification of regulatory miRNA-mRNA networks

Subsequently, an analysis integrating miRNA-mRNA expression data was conducted to build miRNA-mRNA networks of these four miRNAs and differentially expressed genes to understand transcriptome networks regulated by miRNAs in oral cancers. Primarily, target genes of the four miRNAs (miR-140, miR-143, miR-145 and miR-30a) were predicted based on their complementary target sequences using TargetScan, miRDB and miRWalk software. A total of 390 genes were predicted as potential targets of the four identified miRNAs (19-21). Amongst these, 315 genes were found to be differentially expressed in leukoplakia and oral cancer patients based on our transcriptome analysis, TCGA datasets and literature survey (log2 fold change $\geq$ 2; adj p-value <0.01). We constructed a protein-protein interaction network of the identified 315 genes using the STRING database and visualized the results using Cytoscape (22, 23). We used the CytoHubba plugin in Cytoscape to identify the top 16 nodes in the mRNA networks based on the Matthews correlation coefficient (MCC) score and degree of closeness (Supplementary Figure 1 and Supplementary Table 1) (25). Amongst these, the top five hub genes with the highest degree and MCC score were *HIF1A, CDH1, CD44, EGFR,* and *CCND1,* which are considered as vital candidates responsible for oral cancer progression. We also generated interaction networks of these genes using Ingenuity Pathway Analysis (QIAGEN Inc., https://digitalinsights.qiagen.com/IPA) (24). The top three enriched networks generated by IPA comprised of different sets of 35 differentially expressed genes (Z-scores: 41, 36, and 35) that are involved in major effector functions, and cellular and biological mechanisms such as cell cycle progression, developmental disorders, connective tissue disorders, cell death, cellular assembly and organization and cellular function and maintenance (Supplementary Table 2). The other enriched networks were involved in cell-cell signalling, cellular movement, and immune cell tracking. Merging the top five enriched networks depicted that the top nodes of the integrated combined network were in accordance with the top 16 hub genes identified using the CytoHubba plugin of Cytoscape (Figure 6a). Thus, the identified 4-miRNA signature seems to target these 16 hub genes to modulate major effector functions, and cellular and biological mechanisms responsible for oral cancer progression.

Identified miRNA-mRNA regulatory networks modulate signalling pathways responsible for oral cancer progression

To further understand the underlying mechanism by which the four miRNAs modulate various downstream signalling/canonical pathways, we analysed the most significant and common gene networks using Ingenuity Pathway Analysis (QIAGEN Inc., https://digitalinsights.qiagen.com/IPA) (24). Based on the number of genes involved, Z-scores and p-value obtained from the Fischer's exact test, the significance of the association between the genes and the signalling pathways was determined. IPA identified 102 genes eligible for pathway analysis. Statistically significant top canonical pathways included those associated with activation of IL-15 production, activation of MSP-RON signalling in cancer cells, activation of FAK signalling, and inhibition of PTEN signalling pathways. Regulation of the epithelial-mesenchymal transition (EMT) pathway was statistically significant with a higher percentage of overlapping genes compared to those in other pathways; however, its Z-score was not reported in the IPA libraries. Collectively, these pathways were implicated to play a vital role in oral cancer disease progression, aggressiveness, resistance to conventional therapy, and relapse (Figure 6b).

Expression analysis and prognostic relevance of top hub genes

We subsequently compared the differential expression profiles of the 16 hub genes from the current study with those reported in the TCGA oral cancer dataset. A significant upregulation (-log2 (TPM+1) = ±2, $p < 0.01$) was observed in the expression of eight genes (*LOX*, *STAT1*, *PDGFRB*, *JAG1*, *HIF1A*, *GJA1*, *ETS1*, *FLT1*) in cancer patients as compared to their adjacent normal tissue samples. RNA sequencing expression patterns obtained in our study were in accordance with those reported in the TCGA datasets. In addition, overall survival analysis was conducted using the GEPIA2 database (Figure 7) for investigating the prognostic relevance of these 16 hub genes in OSCC patients. The Kaplan-Meier analysis suggested that higher expression levels of CD44 ($p < 0.01$), EGFR ($p < 0.01$), GJA1 ($p < 0.03$), HNF4A ($p < 0.01$), and CCND1 ($p < 0.001$) were significantly associated with poor overall survival of patients (Figure 8). These results clearly indicate the clinical relevance of the 16 identified hub genes controlled by the four miRNAs identified in our study in terms of predicting disease progression and poor prognosis of oral cancer patients.

Discussion:

In this study, we isolated and characterized salivary exosomes derived from leukoplakia and oral cancer patients, conducted a comprehensive analysis of miRNA and mRNA transcriptomes

derived from enriched salivary exosome fractions compared to those derived from patient-matched tumour tissue samples and evaluated their underlying molecular and functional mechanism responsible for disease progression.

Given that, miRNAs differentially expressed in the tumour bulk are more likely to be reflected in the salivary exosomes, we identified a total of 47 miRNAs that were significantly expressed in both salivary exosomes and tumour tissue of pre-cancerous and cancerous patients. Among these, 12 miRNAs were found to be consistently expressed across in all samples as well as the TCGA dataset, which comprises a larger patient cohort, variable sample types, and more stringent statistical parameters. Real time PCR analysis to substantiate these miRNAs in the validation cohort revealed the presence of 7 miRNAs, which included 5 tumour-suppressive (miR-140, miR-143, miR-145, miR-30a, let-7i) and 2 oncogenic miRNAs (miR-21 and miR-423), to be significantly expressed in salivary exosomes and tumour tissue of patient samples as compared to those in tissues of controls. The efficacy of the identified miRNAs as diagnostic signature was determined by constructing ROC curves and conducting PCA. After analysing various combination models, a 4-miRNA signature (miR-30a, miR-140, miR-143 and miR-145) demonstrated significantly higher sensitivity and specificity in effectively differentiating the cancerous and leukoplakia patients from their control counterparts. Moreover, the 4-miRNA combination model demonstrated higher precision in patient derived salivary exosomes, suggesting its potential as liquid biopsy based diagnostic marker for oral cancer detection at an early stage/pre-cancerous stage.

Several biological processes including proliferation, apoptosis, autophagy, cell cycle regulation, and EMT are modulated by miR-30a in various malignancies (26). It has a dual role as a tumour suppressor and an oncogene in a tissue dependent manner (27). In accordance with our findings, Shiah *et al*, reported that tobacco mediated suppression of miR-30a significantly increased the oncogenic activity in oral cancer via hypermethylation (28). Similarly, overexpression of miR-30a has been associated with decreased cell proliferation and increased chemo-sensitivity in oral cancer (29). However, the clinical significance of miR-30a expression or association of ectopic expression of exosomal miR-30a in reversing oral tumorigenesis has not been reported. Thus, our study provides new insights into the dysregulation of miRNAs in the tumour environment.

Studies have reported miR-140, a tumour suppressor encoded in the intronic region of WWP2, to be involved in the modulation of tumour progression, cancer stem cell survival, cell cycle arrest at DNA synthesis stage, regulation of programmed cell death, and invasion (30). In corroboration with our findings, enhanced expression of miR-140 has been reported to significantly correlate with improved overall survival of head and neck cancer patients (31). Also, miR-140-5p suppresses OSCC tumorigenesis by targeting PAK4(32). Moreover, downregulation of exosomal miR-140 has been reported to promote tumorigenesis in breast cancer (33). However, no such studies have been conducted in oral cancer patients using an approach based on liquid biopsy.

The miRNAs miR-143 and miR-145 are transcribed from a single transcript and have been reported to mutually regulate various malignancies such as colorectal (34), breast (35), bladder (36), lung (37), and gastric (38) cancers as crucial tumour suppressors. Our study demonstrated a significant downregulation in the expression of miR-143 and miR-145 in both tumour and salivary exosomal samples. Sun *et al*., and Irani *et al*., have demonstrated that the suppression of miR-143 and miR-145 plays a vital role in promoting cell proliferation, migration, angiogenesis, cell cycle mechanism, cellular glucose metabolism and regulation of cancer stem cell phenotype (39, 40). Downregulation of exosomal miR-145 has been previously identified as an important marker for early detection and risk prediction of oral cancer (41). Exosomal miR-143, which is derived from human mesenchymal stem cells, has demonstrated the potential to suppress prostate cancer progression (42). However, no studies demonstrating the expression and potential of exosomal miR-143 in oral cancers and its potential in suppressing them have been previously reported.

It was imperative to assess the association of the identified miRNA signature with various clinicopathological parameters and patient outcomes considering its high sensitivity and specificity in discriminating between the cancerous patients and their non-cancerous counterparts. Our results demonstrate significant deregulation of tumour- and exosome-derived miR-30a and miR-143 to be clinically correlated with advanced stages and recurrence of the primary malignancy in oral cancer patients. A study by Bufalino et al. revealed that the suppression of the tumour-derived miR-143/145 cluster clinically correlated with progression of the disease, lymph node involvement, and poor survival rates in oral cancer patients (43). To the best of our knowledge, this is the first study reporting clinical association of exosomal miR-143 and miR-30a with disease

progression, poor prognosis, and relapse in oral cancer patients. A noteworthy finding of this study is the association of miR-140, miR-143 and miR-30a with resistance to conventional chemotherapy. Given the small sample size and variability in expression, further validation of this correlation is required with a larger patient cohort. A recent report indicated the crucial role of exosomal-miR-30a in enhancing the sensitivity of conventional chemotherapeutic drugs such as cisplatin and 5-fluorouracil in OSCC resistant cells (44). Collectively, these findings suggested that exosomal expression of the identified miRNA signature has immense clinical utility as a potential biomarker for diagnosis, prognosis, and real-time monitoring of oral cancer patients during treatment.

Dysregulation of miRNAs has a significant impact on transcriptional regulation, cell-fate determination, and alteration of canonical signalling pathways by targeting genes involved in these cellular functions. Changes in miRNA-mRNA molecular interactions disrupt the controlled feedback mechanism resulting in initiation and progression of various solid malignancies including oral cancers (45-49). However, this miRNA-mRNA interplay and the underlying mechanisms responsible for transformation of a normal phenotype to malignancy has been completely unexplored in OSCC. Thus, in this study, we performed a comprehensive paired miRNA and mRNA sequencing assay to identify miRNAs at an early stage of the disease and explore their potential in modulating vital genes responsible for the progression of OSCC.

The integrated miRNA-mRNA network analysis revealed that four miRNAs could potentially target 315 genes that were significantly expressed in both our data and TCGA datasets. An analysis of protein-protein interaction networks of these genes conducted using Cytoscape, identified top 16 hub proteins based on the MCC scores and the highest degree of closeness. Among these, *HIF1α, CDH1, CD44, EGFR,* and *CCND1* were identified as the top hub proteins that have also been reported as driver candidates responsible for oral cancer initiation and progression (50-53). Interestingly, as shown by our results, another study identified miR-143 to have a direct association with CD44v3 and that it could significantly inhibit the ability of migration and invasion of oral tumours (54). Moreover, Shao et al., has demonstrated that miR-145 suppresses cyclin D1 expression, suggesting that miR-145 could effectively reduce OSCC cell growth by regulating the cell cycle mechanism (55). Collectively, these findings substantiate that miR-143 and miR-145 have a role in disease progression by targeting these vital genes and provide a rationale for their

therapeutic applications in OSCC. Amongst the 4-miRNA signature, miR-30a has been reported to enhance chemo-sensitivity by targeting E-cadherin expression in gastric cancers and promote tumorigenesis by regulating HIF-1α expression in renal cancer cells (56). In terms of the miRNA-mRNA association, miR-140 has been reported to suppress the invasive and angiogenic properties by altering EGFR levels in lung cancer (57). However, similar interactions between the identified genes and miR-140 and miR-30a have not been explored in oral cancers.

Further, expression analysis of the 16 hub proteins performed using TCGA datasets, identified eight transcripts (*LOX, STAT1, PDGFRB, JAG1, HIF1A, GJA1, ETS1,* and *FLT1*) with significantly higher expression as compared to that of matched normal adjacent controls. Moreover, high expression of *CD44, EGFR, GJA1, HNF4A,* and *CCND1* showed a good correlation with poor survival of oral cancer patients. Given that the four miRNAs identified in our study interact with the neighbouring microenvironment by targeting these transcripts and interfering with their signalling cascades, it became imperative to understand the underlying mechanism by which these miRNA-mRNA networks modulate various cellular and biological processes involved in oral cancer progression.

These clinically relevant hub proteins comprise of three tyrosine kinase receptors, namely *EGFR, PDGFRB,* and *Flt1 (VEGFR1)* which have been widely reported in oral cancer patients (58-60). Currently, EGFR inhibitors are being used in combination with radiation for the treatment of advanced head and neck tumours (61). These tyrosine kinase receptors have been collectively associated with tumour proliferation, survival, invasion, metastasis, and neo-angiogenesis by dysregulation/activation of the PI3K/AKT/mTOR and MAPK signalling pathways. These classical pathways intricately interconnect and crosstalk with the NF–κB pathway and *CCND1* (identified hub gene), which have the potential to induce Epithelial-mesenchymal transition (EMT) by regulating its transcriptional factor *SNAIL1/2* and modulate G1/S phase of the cell cycle mechanism respectively. Similarly, another hub gene *HNF4α* is reported to interact through the Wnt/β-catenin, NF-κB, and STAT1/3 signalling pathways to increase cell migration and invasion, and decrease apoptosis by modulating the cell cycle mechanism (via *CCND1*) (62). Moreover, *STAT1* has been identified as one of the hub genes that has consistently correlated with poor prognosis and is reported to regulate programmed cell death-ligand 1 (PD-L1), a promising target for immunotherapy in oral cancers (63). Further, CD44, another hub gene, is an established cancer

stem cell marker for various malignancies, particularly for head and neck cancers. Altered expression of CD44 has been correlated with disease progression, disease aggressiveness, relapse, and resistance to conventional therapy via activation of the same classical pathways of the tyrosine kinase receptors (64, 65). However, the interaction of CD44 with the NANOG, SOX2/9, OCT4, and ERK1/2 signalling pathway is attributed to its enhanced stemness characteristics (66). Similar to our results, Gai, et al. (2018) showed that higher presence of CD44 on the surface of salivary exosomes of cancer patients was correlated with poor patient outcomes (67). Accumulating evidence suggests that the NOTCH signalling pathway is one of the most frequently altered pathways in head and neck squamous cell carcinomas (HNSCC) (68). Given the importance of NOTCH signalling in regulating tumour cell behaviour, NOTCH-targeted strategies are currently being tested in preclinical and clinical settings. Interestingly, proteins of the Notch receptor family bind to the ligand JAG1 and increase the recruitment of *HIF-1α* to the *LOX* promoter region. Elevated LOX expression activates the SNAIL protein, which in turn results in migration, invasion, and EMT in cancer (69). In our study, *HIF-1α, LOX*, and *JAG1* have been identified as hub genes, thereby suggesting an inevitable role of this underlying mechanism in oral cancer progression. ETS-1 is a family of transcriptional factors that reactivate the ERK signalling pathway via the AKT/ETS-1 positive feedback loop and are responsible for tumorigenesis, cell death, resistance to EGFR tyrosine kinase inhibitors and inducing stemness properties (70) (Figure 9).

This study identifies a robust 4-miRNA signature that can be utilized for the early detection of oral cancers using the liquid biopsy approach. Further, these four miRNAs target signalling cascades, to modulate various biological processes such as tumour progression, survival, invasion, metastasis, neo-angiogenesis, resistance to therapy and recurrence of the disease. These biological processes are associated with hallmarks of cancer such as sustaining proliferative signalling, resisting cell death, activating invasion and metastasis, inducing angiogenesis, evading growth suppressors, and enabling replicative immortality. Therefore, our findings reveal the immense diagnostic and prognostic potential of these four miRNAs and uncover the plausible mechanism by which dysregulation of these tumour suppressive miRNAs could effectively convert a normal epithelial phenotype into oral carcinoma.

Data availability statement: RNA sequencing data will be deposited in the Gene Expression Omnibus (GEO) database (Accession number) and will be available concomitant with publication. Other data that support the findings are available upon request from the corresponding authors.

Geolocation information: Biological and Life Sciences, School of Arts and Sciences, Ahmedabad University, Commerce Six roads, Navrangpura, Ahmedabad – 380009, Gujarat, India.

Acknowledgement: The authors are grateful to Prof. Rakesh Rawal, Gujarat University for kindly providing access to the NanoSight LM10. We acknowledge, Dept. of Bioinformatics, Gujarat University for providing computational resources for data analysis. This facility was funded by the Gujarat State Biotechnology Mission and Gujarat Council on Science and Technology, Government of Gujarat.

Funding: This work was supported by The Gujarat State Biotechnology Mission- Financial Assistance Programme under the Grant Number: 0P3C9K; and Ahmedabad University Start-up grant under the Grant Number AU/SUG/SAS/BLS/2018-19/04-Vivek Tanavde_03.22. Financial assistance to SP from the National Postdoctoral Fellowship Program by Department of Science and Technology – Science and Engineering Research Board, Government of India (PDF/2017/001912) is gratefully acknowledged.

Disclosure of interest: The authors declare no competing interests.

Ethics approval statement: All procedures of human samples were conducted after the approval of the HCG Cancer Centre's Ethics Committee (ECR/92/Inst/GJ/2013/RR-16) for human subject research. This study also complied with the guidelines set forth by the Declaration of Helsinki (2008). All patients provided written informed consent for their participation in the study and their identities have been anonymised.

Author contributions: KP and DM: provided clinical samples and clinical assistance; SP, KP, KS and VT: conceived and designed the experiments; SP and AP: conducted the experiments; SP, AP, and KS: performed bioinformatic analysis; SP and AP: performed data analysis; SP, AP, and VT: wrote the manuscript; SP, KS, and VT: edited the manuscript.

References:

1. Malik UU, Zarina S, Pennington SR. Oral squamous cell carcinoma: Key clinical questions, biomarker discovery, and the role of proteomics. Archives of oral biology. 2016;63:53-65.
2. Ge R, Tan E, Sharghi-Namini S, Asada HH. Exosomes in Cancer Microenvironment and Beyond: have we Overlooked these Extracellular Messengers? Cancer Microenvironment. 2012;5(3):323-32.
3. Kourembanas S. Exosomes: vehicles of intercellular signaling, biomarkers, and vectors of cell therapy. Annual review of physiology. 2015;77:13-27.
4. Osaki M, Okada F. Exosomes and their role in cancer progression. Yonago acta medica. 2019;62(2):182-90.
5. Patel A, Patel S, Patel P, Tanavde V. Saliva Based Liquid Biopsies in Head and Neck Cancer: How Far Are We From the Clinic? Frontiers in Oncology. 2022;12:828434-.
6. Li C, Zhou T, Chen J, Li R, Chen H, Luo S, et al. The role of Exosomal miRNAs in cancer. Journal of translational medicine. 2022;20(1):1-15.
7. Sun Z, Shi K, Yang S, Liu J, Zhou Q, Wang G, et al. Effect of exosomal miRNA on cancer biology and clinical applications. Molecular cancer. 2018;17(1):1-19.
8. Andrés-León E, Cases I, Alonso S, Rojas AM. Novel miRNA-mRNA interactions conserved in essential cancer pathways. Scientific reports. 2017;7(1):1-13.
9. Ventura TMO, Ribeiro NR, Taira EA, de Lima Leite A, Dionizio A, Rubira CMF, et al. Radiotherapy changes the salivary proteome in head and neck cancer patients: evaluation before, during, and after treatment. Clinical Oral Investigations. 2022;26(1):225-58.
10. Prendergast EN, de Souza Fonseca MA, Dezem FS, Lester J, Karlan BY, Noushmehr H, et al. Optimizing exosomal RNA isolation for RNA-Seq analyses of archival sera specimens. PLoS One. 2018;13(5):e0196913.
11. Andrews S. FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom; 2010.
12. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetjournal. 2011;17(1):10.

13. Friedländer MR, Mackowiak SD, Li N, Chen W, Rajewsky N. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. Nucleic acids research. 2012;40(1):37-52.

14. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15-21.

15. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nature protocols. 2016;11(9):1650-67.

16. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30(7):923-30.

17. Arbuckle JL. IBM SPSS Amos 20 user's guide. Amos Development Corporation, SPSS Inc. 2011:226-9.

18. Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K. Cluster: cluster analysis basics and extensions. R package version. 2012;1(2):56.

19. Wang X. miRDB: a microRNA target prediction and functional annotation database with a wiki interface. Rna. 2008;14(6):1012-7.

20. Agarwal V, Bell GW, Nam J-W, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. elife. 2015;4:e05005.

21. Dweep H, Sticht C, Pandey P, Gretz N. miRWalk–database: prediction of possible miRNA binding sites by "walking" the genes of three genomes. Journal of biomedical informatics. 2011;44(5):839-47.

22. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al. The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. Nucleic acids research. 2016:gkw937.

23. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research. 2003;13(11):2498-504.

24. Krämer A, Green J, Pollard Jr J, Tugendreich S. Causal analysis approaches in ingenuity pathway analysis. Bioinformatics. 2014;30(4):523-30.

25. Chin C-H, Chen S-H, Wu H-H, Ho C-W, Ko M-T, Lin C-Y. cytoHubba: identifying hub objects and sub-networks from complex interactome. BMC systems biology. 2014;8(4):1-7.

26. Jiang L-h, Zhang H-d, Tang J-h. MiR-30a: A novel biomarker and potential therapeutic target for cancer. Journal of oncology. 2018;2018.

27. Wang C, Yang X, Chen Y, Chen L. The versatile role of microRNA-30a in human cancer. Cellular Physiology and Biochemistry. 2017;41(4):1616-32.

28. Shiah S-G, Hsiao J-R, Chang H-J, Hsu Y-M, Wu G-H, Peng H-Y, et al. MiR-30a and miR-379 modulate retinoic acid pathway by targeting DNA methyltransferase 3B in oral cancer. Journal of biomedical science. 2020;27(1):1-13.

29. Kawahara K, Nagata M, Yoshida R, Hirosue A, Tanaka T, Matsuoka Y, et al. miR-30a attenuates drug sensitivity to 5-FU by modulating cell proliferation possibly by downregulating cyclin E2 in oral squamous cell carcinoma. Biochemistry and biophysics reports. 2021;28:101114.

30. Green D, Dalmay T, Fraser WD. Role of miR-140 in embryonic bone development and cancer. Clinical Science. 2015;129(10):863-73.

31. Zheng M, Liu J, Meng C, Tang K, Liao J. Prognostic and clinicopathological importance of microRNA-140 expression in cancer patients: a meta-analysis. World Journal of Surgical Oncology. 2021;19(1):1-9.

32. Peng M, Pang C. MicroRNA-140-5p inhibits the tumorigenesis of oral squamous cell carcinoma by targeting p21-activated kinase 4. Cell Biology International. 2020;44(1):145-54.

33. Gernapudi R, Yao Y, Zhang Y, Wolfson B, Roy S, Duru N, et al. Targeting exosomes from preadipocytes inhibits preadipocyte to cancer stem cell signaling in early-stage breast cancer. Breast cancer research and treatment. 2015;150(3):685-95.

34. Pagliuca A, Valvo C, Fabrizi E, Di Martino S, Biffoni M, Runci D, et al. Analysis of the combined action of miR-143 and miR-145 on oncogenic pathways in colorectal cancer cells reveals a coordinate program of gene repression. Oncogene. 2013;32(40):4806-13.

35. Yan X, Chen X, Liang H, Deng T, Chen W, Zhang S, et al. miR-143 and miR-145 synergistically regulate ERBB3 to suppress cell proliferation and invasion in breast cancer. Molecular cancer. 2014;13(1):1-14.

36. Villadsen S, Bramsen J, Ostenfeld M, Wiklund E, Fristrup N, Gao S, et al. The miR-143/-145 cluster regulates plasminogen activator inhibitor-1 in bladder cancer. British journal of cancer. 2012;106(2):366-74.

37. Dimitrova N, Gocheva V, Bhutkar A, Resnick R, Jong RM, Miller KM, et al. Stromal expression of miR-143/145 promotes neoangiogenesis in lung cancer development. Cancer discovery. 2016;6(2):188-201.

38. Lei C, Du F, Sun L, Li T, Li T, Min Y, et al. miR-143 and miR-145 inhibit gastric cancer cell migration and metastasis by suppressing MYO6. Cell death & disease. 2017;8(10):e3101.

39. Sun X, Zhang L. MicroRNA-143 suppresses oral squamous cell carcinoma cell growth, invasion and glucose metabolism through targeting hexokinase 2. Bioscience Reports. 2017;37(3).

40. Irani S, Shokri G. The role of miR-143, miR-145, and miR-590 in expression levels of CD44 and vascular endothelial cadherin in oral squamous cell carcinoma. Middle East Journal of Cancer. 2019;10(3):194-204.

41. Hang W, Feng Y, Sang Z, Yang Y, Zhu Y, Huang Q, et al. Downregulation of miR-145-5p in cancer cells and their derived exosomes may contribute to the development of ovarian cancer by targeting CT. International journal of molecular medicine. 2019;43(1):256-66.

42. Che Y, Shi X, Shi Y, Jiang X, Ai Q, Shi Y, et al. Exosomes derived from miR-143-overexpressing MSCs inhibit cell migration and invasion in human prostate cancer by downregulating TFF3. Molecular Therapy-Nucleic Acids. 2019;18:232-44.

43. Bufalino A, Cervigne NK, De Oliveira CE, Fonseca FP, Rodrigues PC, Macedo CCS, et al. Low miR-143/miR-145 cluster levels induce activin A overexpression in oral squamous cell carcinomas, which contributes to poor prognosis. PloS one. 2015;10(8):e0136599.

44. Kulkarni B, Gondaliya P, Kirave P, Rawal R, Jain A, Garg R, et al. Exosome-mediated delivery of miR-30a sensitize cisplatin-resistant variant of oral squamous carcinoma cells via modulating Beclin1 and Bcl2. Oncotarget. 2020;11(20):1832.

45. Sun J, Gong X, Purow B, Zhao Z. Uncovering microRNA and transcription factor mediated regulatory networks in glioblastoma. PLoS computational biology. 2012;8(7):e1002488.

46. Ying H, Lv J, Ying T, Li J, Yang Q, Ma Y. microRNA and transcription factor mediated regulatory network for ovarian cancer. Tumor Biology. 2013;34(5):3219-25.

47. Murray MJ, Saini HK, van Dongen S, Palmer RD, Muralidhar B, Pett MR, et al. The two most common histological subtypes of malignant germ cell tumour are distinguished by global microRNA profiles, associated with differential transcription factor expression. Molecular cancer. 2010;9(1):1-14.

48. Riba A, Bosia C, El Baroudi M, Ollino L, Caselle M. A combination of transcriptional and microRNA regulation improves the stability of the relative concentrations of target genes. PLoS computational biology. 2014;10(2):e1003490.

49. Zheng X, Wu K, Liao S, Pan Y, Sun Y, Chen X, et al. MicroRNA-transcription factor network analysis reveals miRNAs cooperatively suppress RORA in oral squamous cell carcinoma. Oncogenesis. 2018;7(10):1-18.

50. Chien HT, Cheng SD, Liao CT, Wang HM, Huang SF. Amplification of the EGFR and CCND1 Are Coordinated and Play Important Roles in the Progression of Oral Squamous Cell Carcinomas. Cancers. 2019;11(6):760.

51. Frohwitter G, Buerger H, Korsching E, van Diest PJ, Kleinheinz J, Fillies T. Site-specific gene expression patterns in oral cancer. Head & face medicine. 2017;13(1):1-9.

52. Wen G, Wang H, Zhong Z. Associations of RASSF1A, RARβ, and CDH1 promoter hypermethylation with oral cancer risk: A PRISMA-compliant meta-analysis. Medicine. 2018;97(11).

53. Judd NP, Winkler AE, Murillo-Sauca O, Brotman JJ, Law JH, Lewis JS, et al. ERK1/2 regulation of CD44 modulates oral cancer aggressiveness. Cancer research. 2012;72(1):365-74.

54. Xu P, Li Y, Yang S, Yang H, Tang J, Li M. Micro-ribonucleic acid 143 (MiR-143) inhibits oral squamous cell carcinoma (OSCC) cell migration and invasion by downregulation of phospho-c-Met through targeting CD44 v3. Oral surgery, oral medicine, oral pathology and oral radiology. 2015;120(1):43-51.

55. Shao Y, Qu Y, Dang S, Yao B, Ji M. MiR-145 inhibits oral squamous cell carcinoma (OSCC) cell growth by targeting c-Myc and Cdk6. Cancer cell international. 2013;13(1):1-9.

56. Wang L, Zhang X, Zhang X, Chu J. MiR-30a increases cisplatin sensitivity of gastric cancer cells through suppressing epithelial-to-mesenchymal transition (EMT). Eur Rev Med Pharmacol Sci. 2016;20(9):1733-9.

57. Flamini V, Dudley WGJ, Cui Y. Distinct mechanisms by which two forms of miR-140 suppress the malignant properties of lung cancer cells. Oncotarget. 2018;9(92):36474.

58. Huang S-F, Cheng S-D, Chien H-T, Liao C-T, Chen I-H, Wang H-M, et al. Relationship between epidermal growth factor receptor gene copy number and protein expression in oral cavity squamous cell carcinoma. Oral oncology. 2012;48(1):67-72.

59. Lin L-H, Lin J-S, Yang C-C, Cheng H-W, Chang K-W, Liu C-J. Overexpression of platelet-derived growth factor and its receptor are correlated with oral tumorigenesis and poor prognosis in oral squamous cell carcinoma. International journal of molecular sciences. 2020;21(7):2360.

60. Subarnbhesaj A, Miyauchi M, Chanbora C, Mikuriya A, Nguyen PT, Furusho H, et al. Roles of VEGF-Flt-1 signaling in malignant behaviors of oral squamous cell carcinoma. PLoS One. 2017;12(11):e0187092.

61. Cuneo KC, Nyati MK, Ray D, Lawrence TS. EGFR targeted therapies and radiation: Optimizing efficacy by appropriate drug scheduling and patient selection. Pharmacology & therapeutics. 2015;154:67-77.

62. Lv D-D, Zhou L-Y, Tang H. Hepatocyte nuclear factor 4α and cancer-related cell signaling pathways: a promising insight into cancer treatment. Experimental & Molecular Medicine. 2021;53(1):8-18.

63. Cui B, Chen J, Luo M, Liu Y, Chen H, Lü D, et al. PKD3 promotes metastasis and growth of oral squamous cell carcinoma through positive feedback regulation with PD-L1 and activation of ERK-STAT1/3-EMT signalling. International journal of oral science. 2021;13(1):1-10.

64. Patel S, Rawal R. Role of miRNA dynamics and cytokine profile in governing CD44v6/Nanog/PTEN axis in oral cancer: modulating the master regulators. Tumor Biology. 2016;37(11):14565-75.

65. Patel S, Shah K, Mirza S, Shah K, Rawal R. Circulating tumor stem like cells in oral squamous cell carcinoma: An unresolved paradox. Oral Oncology. 2016;62:139-46.

66. Huang C, Yoon C, Zhou X-H, Zhou Y-C, Zhou W-W, Liu H, et al. ERK1/2-Nanog signaling pathway enhances CD44 (+) cancer stem-like cell phenotypes and epithelial-to-mesenchymal transition in head and neck squamous cell carcinomas. Cell death & disease. 2020;11(4):1-14.

67. Gai C, Camussi F, Broccoletti R, Gambino A, Cabras M, Molinaro L, et al. Salivary extracellular vesicle-associated miRNAs as potential biomarkers in oral squamous cell carcinoma. BMC cancer. 2018;18(1):1-11.

68. Sun W, Gaykalova DA, Ochs MF, Mambo E, Arnaoutakis D, Liu Y, et al. Activation of the NOTCH pathway in head and neck cancer. Cancer research. 2014;74(4):1091-104.

69. Maggiolini M, Musti AM, De Francesco EM. Crosstalk between Notch, HIF-1α and GPER in Breast Cancer EMT. International Journal of Molecular Sciences. 2018;19(7).

70. Dittmer J, editor The role of the transcription factor Ets1 in carcinoma. Seminars in cancer biology; 2015: Elsevier.

Figure Legends:

Figure 1: Overview of the study design. This figure explains the design of this study. It shows the sequencing and data analysis pipeline.

Figure 2: Characterisation of exosomes based on size, concentration, and expression of tetraspanin. (a-b) Estimation of size and concentration of salivary exosomes from healthy volunteers and OSCC patients using Nanoparticle Tracking Analysis (NTA). The x-axis indicates the size distribution of particles and the y-axis shows the relative concentration (particles/ml) in NTA; **(c-d)** Representative TEM images of exosomes from control volunteer - and OSCC patient, (Scale: 200nm); **(e-f)** Representative plots of salivary exosomes stained with anti-CD63-Alexa fluor 488, anti-CD81-PE, and anti-CD9-APC antibodies analysed by flow cytometry.

Figure 3: Differential expression of miRNAs in OSCC patients compared to healthy controls. (a) Venn diagram representing the forty-seven common miRNAs that were differentially expressed in tissue and salivary exosomes derived from OSCC and leucoplakia patients at nominal adj $p < 0.05$. **(b)** Volcano plot depicting the twelve significantly expressed miRNAs in both our and TCGA datasets ($p < 0.001$). **(c)** Box plots depicting the differential expression patterns of seven miRNAs in larger cohorts of tumour tissue and salivary exosomes of OSCC and leucoplakia patients using real-time PCR (***$p \leq 0.001$). U6 small RNA was used as internal control, and the relative mRNA levels were analysed using the ddCt method. Error bars represent mean ± SD of three independent experiments.

Figure 4: 4 miRNA panel can differentiate between cancerous and non-cancerous samples with very good sensitivity and specificity. (a) In tissue cohort, the combined measure of sensitivity and specificity of 7 miRNAs (miR-140, miR-143, miR-145, miR-30a, miR-21, miR-423 and let7i) and 4 miRNAs (miR-140, miR-143, miR-145, miR-30a) was represented by AUC at 95% CI, p-value<0.001) which equals to 0.80 (0.74–0.86, $p < 0.0001$) and 0.92 (0.87–0.97, $p < 0.0001$) respectively. **(b)** In salivary exosome cohort, the combined AUC of 7 miRNAs and 4 miRNAs at 95% CI, p-value<0.001) was 0.92 (0.88–0.96, $p < 0.0001$), and 0.99 (0.98–1.0, $p < 0.0001$), respectively. Diagonal reference line acts as a performance measure of the diagnostic test. Note: AUC, area under the curve, CI, confidence interval. **(c)** Representative images are 3D plots of PCA performed using expression of four significantly expressed miRNAs. These plots suggest that expression of these 4 miRNAs can distinguish OSCC tissue samples and salivary exosomes from their normal counterparts with 81.12% variance and 89.98% variance respectively.

Figure 5: Association of identified four-miRNA signature with various clinicopathological parameters. Representative box plots depict relative expression patterns of miRNAs in early

stages (Stage I/II) vs late stages (Stage III/IV) of OSCC patient derived **(a)** tissue samples and **(b)** salivary exosomal samples. **(c)** Representative box plots depict relative expression patterns of tumour-derived miRNAs in **(d)** recurrent vs non-recurrent OSCC patients and **(e)** chemoresistant vs non-chemo-resistant OSCC patients. The expression levels of miRNAs were estimated using real-time PCR. The data were normalised with U6 values, and the relative miRNA levels were analysed using the ddCt method. Error bars represent mean ± SD of three independent experiments (*$p \leq 0.05$; **$p \leq 0.05$; ***$p \leq 0.001$).

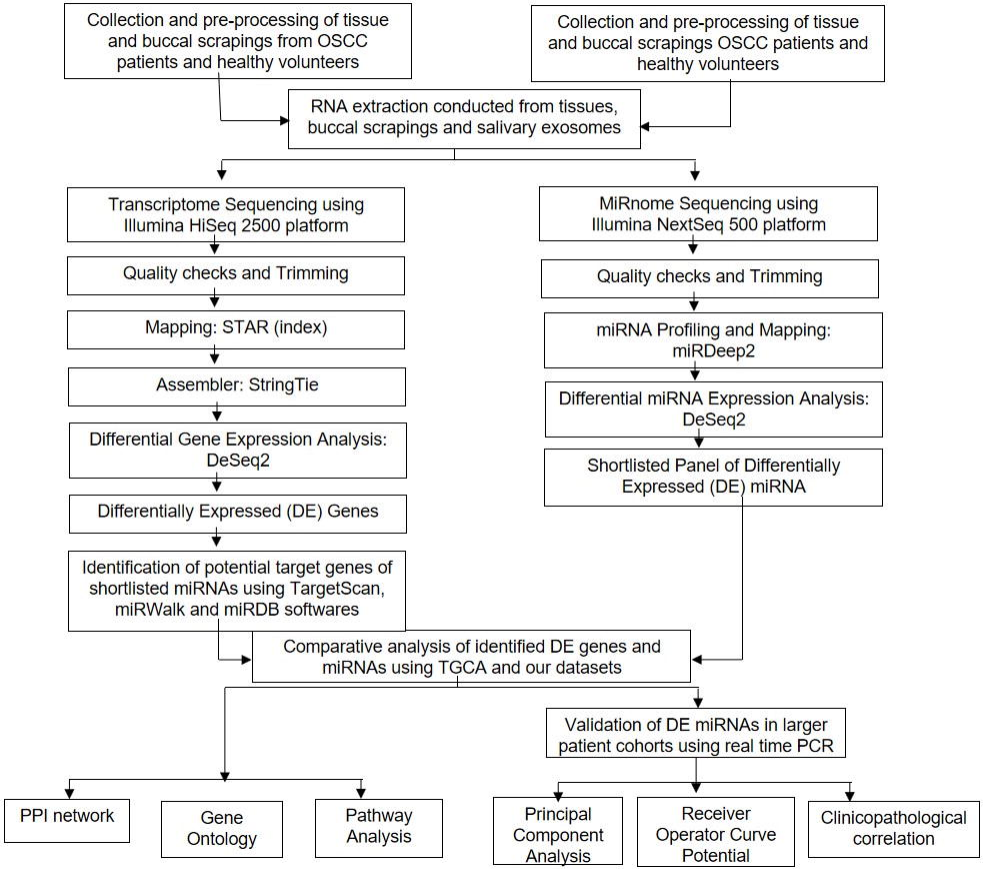
Figure 6: miRNA:mRNA gene network and canonical pathways identified by Ingenuity Pathway Analysis (IPA): (a) The miRNA:mRNA gene network from experimentally observed miRNA–mRNA interactions of four miRNAs and 16 hub genes. Each network displays the genes/gene products as nodes (different shapes representing the functional classes of gene products) and the biological relationships between the nodes as lines. Full lines represent direct interactions and broken lines, indirect interactions. Molecules shown in red represent the most centrally located nodes. Orange colored nodes depict the key nodes that are responsible for deregulating the major signalling pathways in tumorigenesis. The miRNAs are represented as yellow molecules that govern the regulatory network. **(b)** Differentially expressed canonical pathways targeted by 16 hub genes. Higher -log Benjamini-Hochberg (B-H) p value corresponds to the more significant pathway. Z score is the activation score for a pathway. The Z score reflects the activation (positive-orange colour) or inhibition (negative-blue colour) of a pathway due to the changes in the expression of genes involved in those pathways or functions.

Figure 7: Differential expression of top hub genes based on HNSCC TCGA dataset using the GEPIA platform. Differential expression of **(a)** LOX **(b)** STAT1 **(C)** PDGFRB **(d)** JAG1 **(e)** HIF1α **(f)** GJA1 **(g)** EST1 **(h)** FLT1 that were found to be significantly expressed in the HNSCC TCGA dataset as compared to their representative controls ($p < 0.01$), -log2(TPM+1).

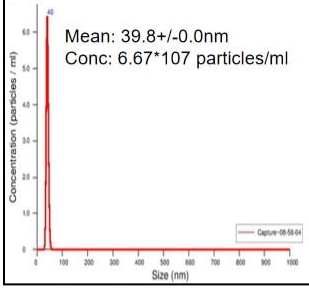
Figure 8: Overall survival analysis of hub genes performed using the GEPIA platform. Overall survival analyses performed using the GEPIA platform where patients with hub genes **(a)** CD34 **(b)** STAT1 **(c)** CD44 **(d)** EGFR **(e)** GJA1 **(f)** HNF4A **(g)** CCND1 expression above the median are indicated by red lines, and patients with hub gene expression below the median are indicated by black lines. Log-rank $p < 0.05$ was considered to indicate a statistically significant difference.

Figure 9: Schematic representation of putative signalling pathways regulated by the identified miRNA-mRNA networks responsible for OSCC progression. The processes in green boxes are associated with hallmarks of cancer.

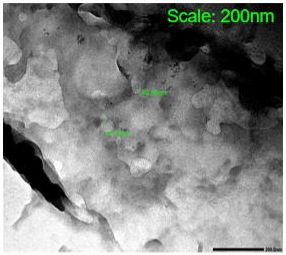
Supplementary Figure 1: Top 16 hub genes based on the maximal clique centrality (MCC). Protein–Protein interaction network generated in Cytoscape using STRING plugin. The hub gene analysis was performed by CytoHubba. The blue coloured nodes represent the 16 nodes with the highest MCC score.


Collection and pre-processing of tissue and buccal scrapings from OSCC patients and healthy volunteers
Collection and pre-processing of tissue and buccal scrapings OSCC patients and healthy volunteers
RNA extraction conducted from tissues, buccal scrapings and salivary exosomes
Transcriptome Sequencing using Illumina HiSeq 2500 platform
Quality checks and Trimming
Mapping: STAR (index)
Assembler: StringTie
Differential Gene Expression Analysis: DeSeq2
Differentially Expressed (DE) Genes
Identification of potential target genes of shortlisted miRNAs using TargetScan, miRWalk and miRDB softwares
MiRnome Sequencing using Illumina NextSeq 500 platform
Quality checks and Trimming
miRNA Profiling and Mapping: miRDeep2
Differential miRNA Expression Analysis: DeSeq2
Shortlisted Panel of Differentially Expressed (DE) miRNA
Comparative analysis of identified DE genes and miRNAs using TGCA and our datasets
Validation of DE miRNAs in larger patient cohorts using real time PCR
PPI network
Gene Ontology
Pathway Analysis
Principal Component Analysis
Receiver Operator Curve Potential
Clinicopathological correlation

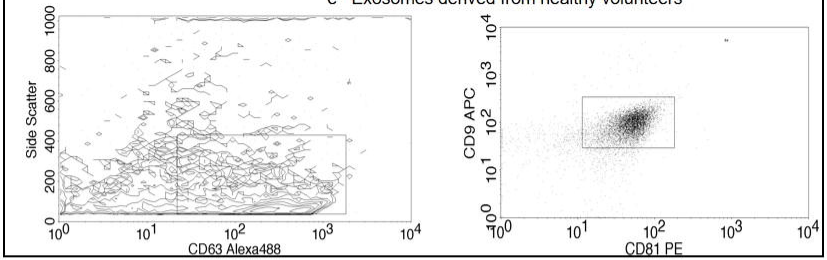
a Exosomes derived from healthy volunteers



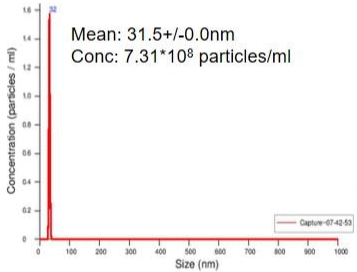
c Exosomes derived from healthy volunteers



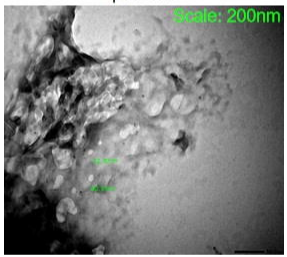
e Exosomes derived from healthy volunteers



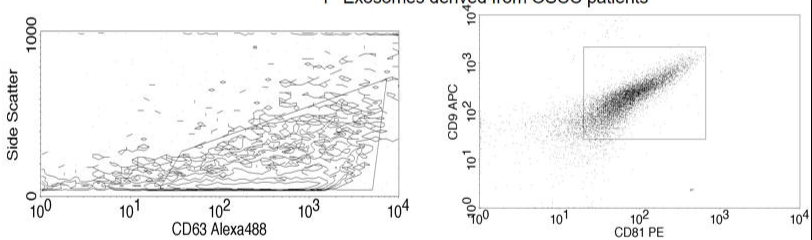
b Exosomes derived from OSCC patients



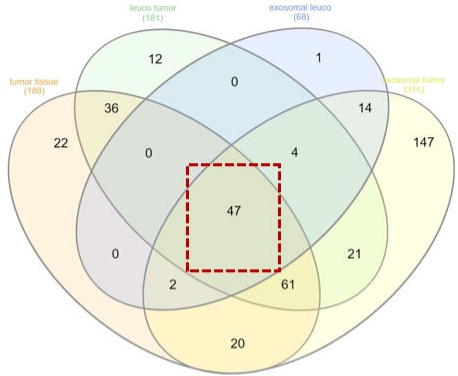
d Exosomes derived from OSCC patients



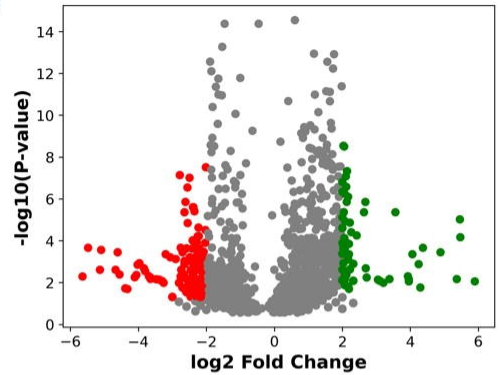
f Exosomes derived from OSCC patients



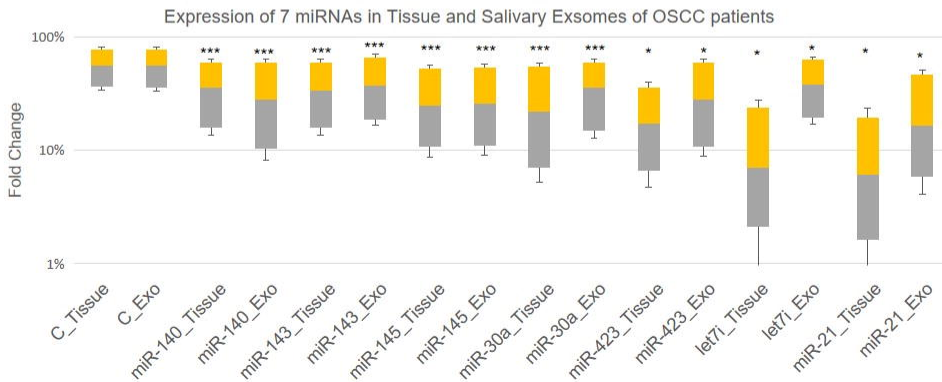
a


leuco tumor
(181)
exosomal leuco
(68)
tumor tissue
(188)
exosomal tumor
(316)
12
1
0
36
14
22
0
4
147
47
0
21
2
61
20

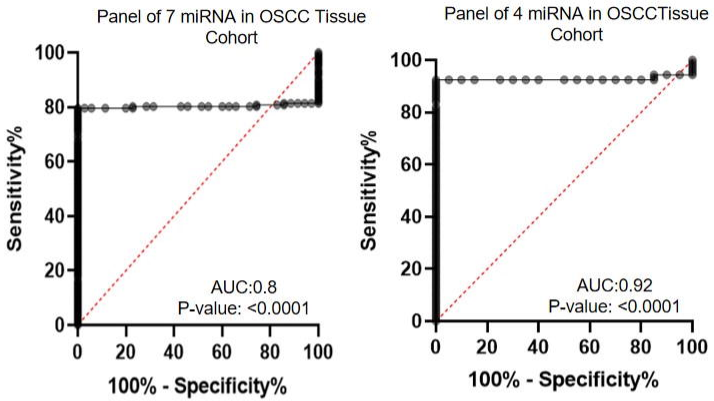
b


-log10(P-value)
log2 Fold Change

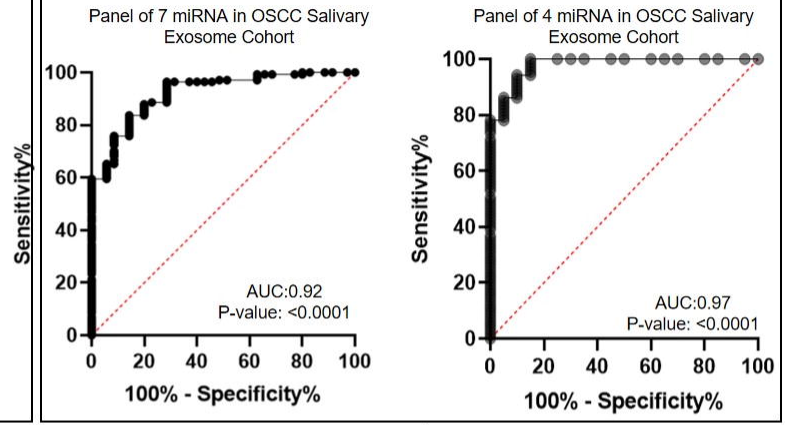
c


Expression of 7 miRNAs in Tissue and Salivary Exsomes of OSCC patients
Fold Change
100%
10%
1%
C_Tissue
C_Exo
miR-140_Tissue
miR-140_Exo
miR-143_Tissue
miR-143_Exo
miR-145_Tissue
miR-145_Exo
miR-30a_Tissue
miR-30a_Exo
miR-423_Tissue
miR-423_Exo
let7i_Tissue
let7i_Exo
miR-21_Tissue
miR-21_Exo

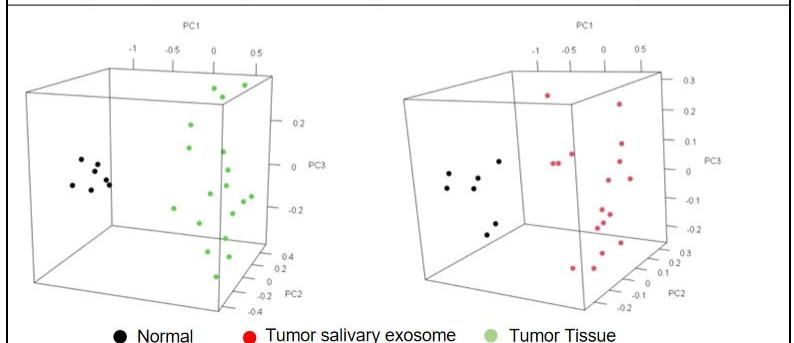
a Sensitivity And Specifivity Of miRNA Panel In OSCC Tissue Cohort

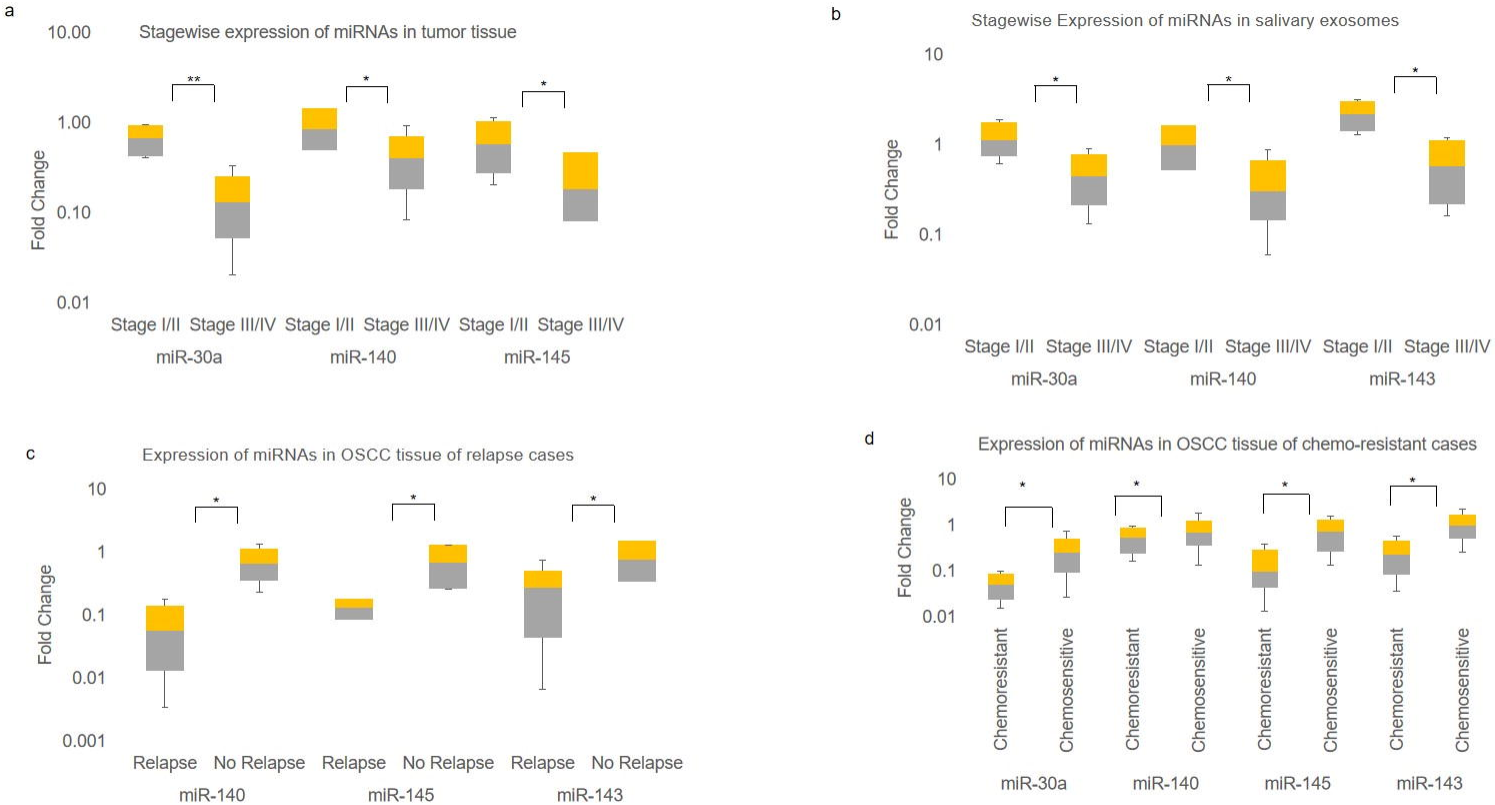


b Sensitivity And Specificity Of miRNA Panel In OSCC Salivary Exosome Cohort

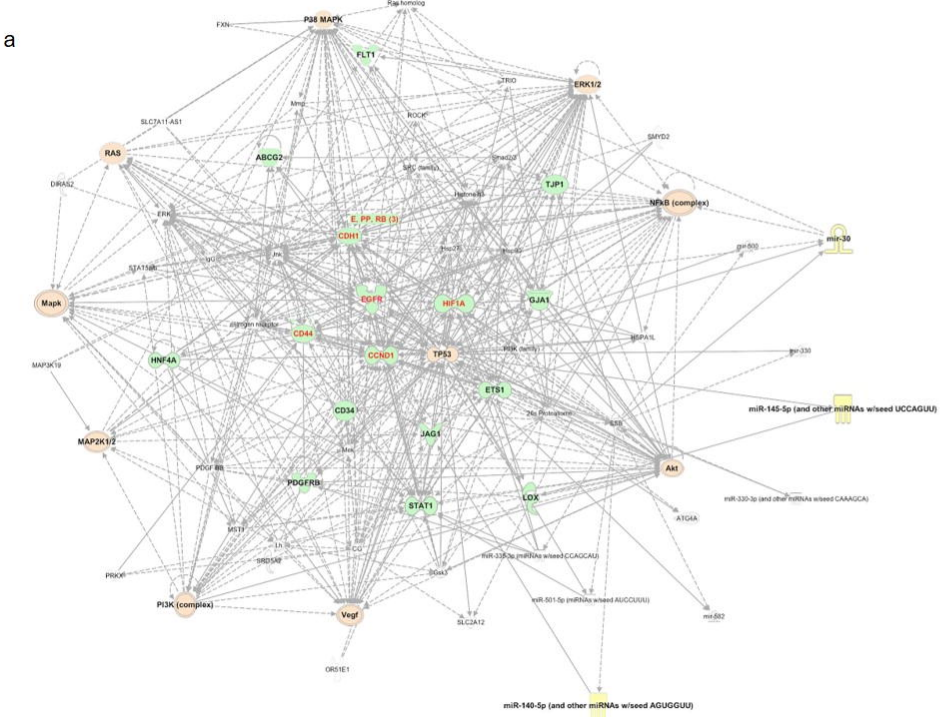


c Principal Component Analysis of 4 miRNA panel in OSCC tissue and salivary exosome cohort

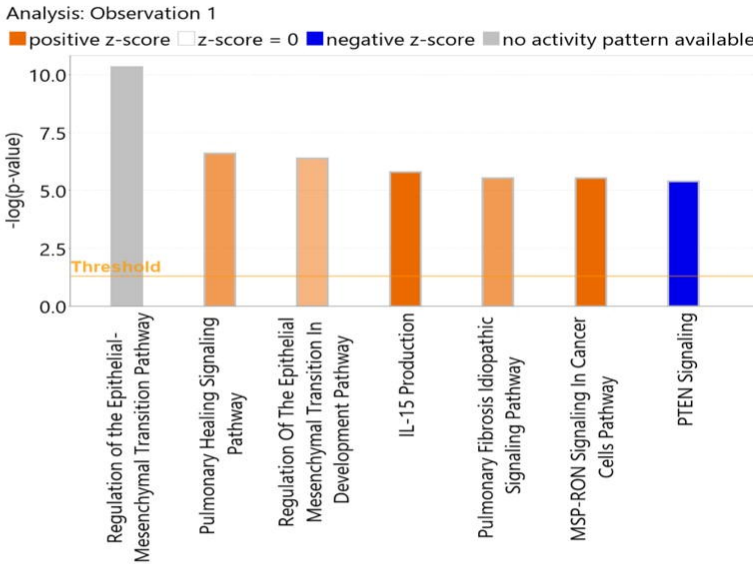



a
Stagewise expression of miRNAs in tumor tissue
Fold Change
10.00
1.00
0.10
0.01
**
*
*
Stage I/II
Stage III/IV
Stage I/II
Stage III/IV
Stage I/II
Stage III/IV
miR-30a
miR-140
miR-145
b
Stagewise Expression of miRNAs in salivary exosomes
Fold Change
10
1
0.1
0.01
*
*
*
Stage I/II
Stage III/IV
Stage I/II
Stage III/IV
Stage I/II
Stage III/IV
miR-30a
miR-140
miR-143
c
Expression of miRNAs in OSCC tissue of relapse cases
Fold Change
10
1
0.1
0.01
0.001
*
*
*
Relapse
No Relapse
Relapse
No Relapse
Relapse
No Relapse
miR-140
miR-145
miR-143
d
Expression of miRNAs in OSCC tissue of chemo-resistant cases
Fold Change
10
1
0.1
0.01
*
*
*
*
Chemoresistant
Chemosensitive
Chemoresistant
Chemosensitive
Chemoresistant
Chemosensitive
Chemoresistant
Chemosensitive
miR-30a
miR-140
miR-145
miR-143

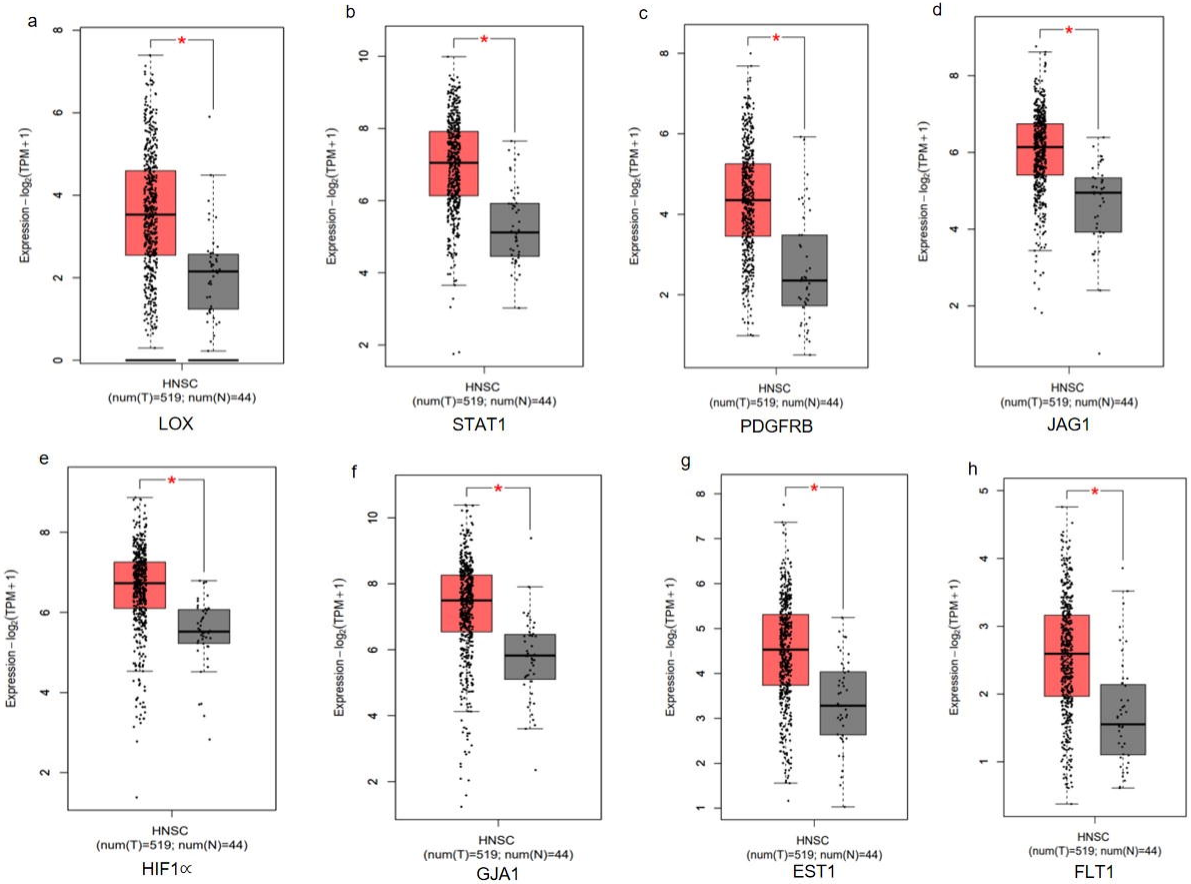
a

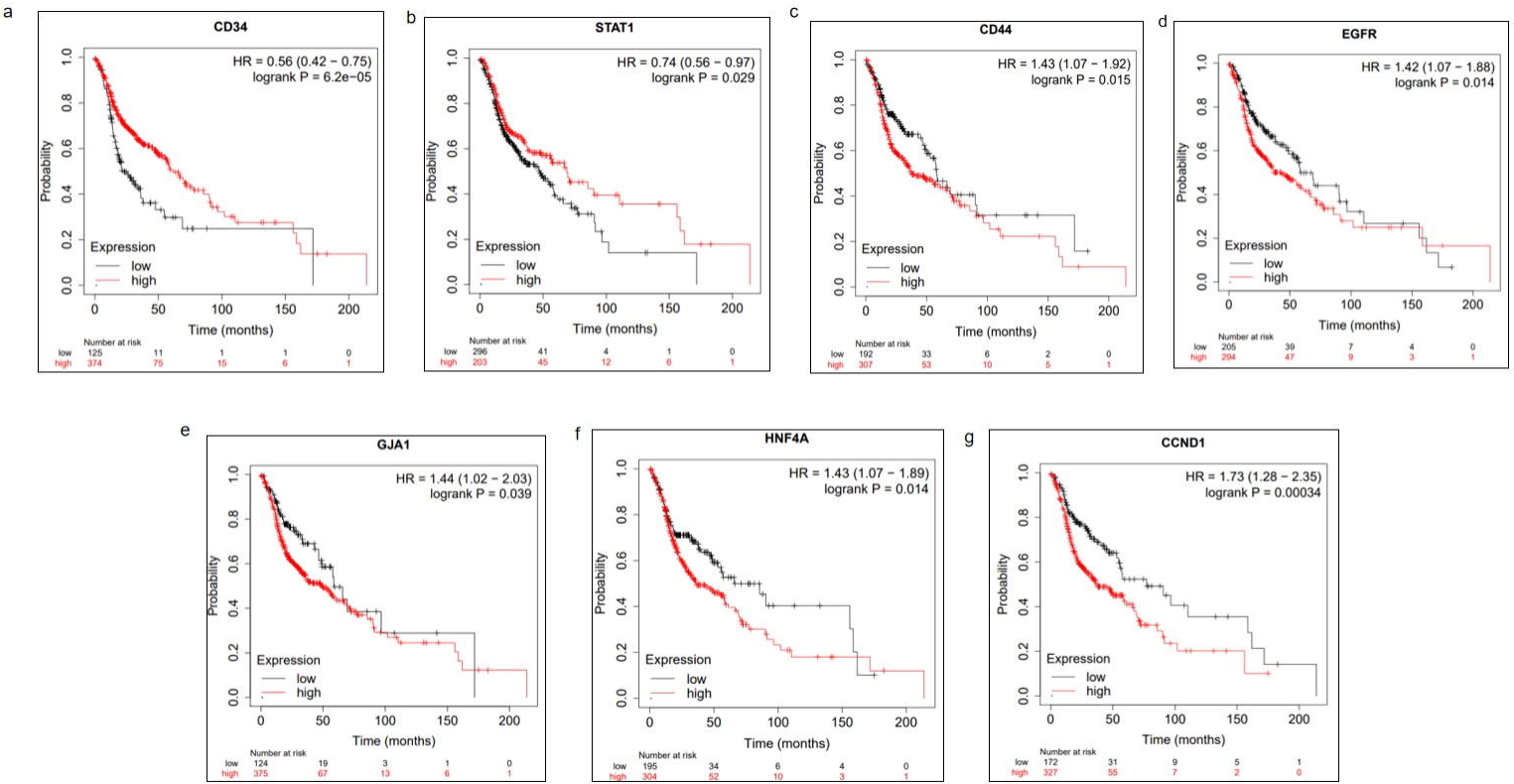


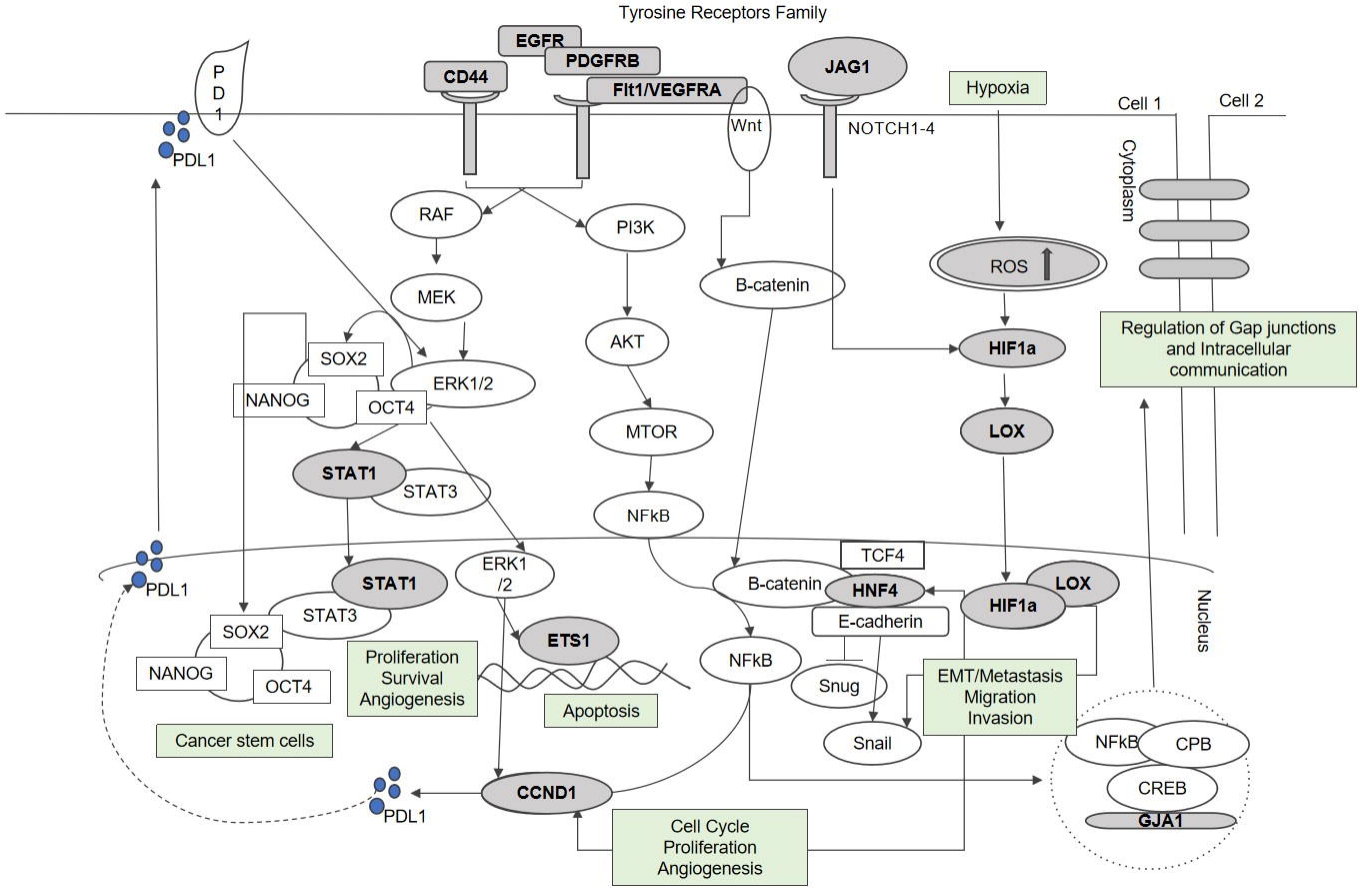
b



© 2000-2022 QIAGEN. All rights reserved.


a
Expression – log2(TPM+1)
HNSC
(num(T)=519; num(N)=44)
LOX
b
Expression – log2(TPM+1)
HNSC
(num(T)=519; num(N)=44)
STAT1
c
Expression – log2(TPM+1)
HNSC
(num(T)=519; num(N)=44)
PDGFRB
d
Expression – log2(TPM+1)
HNSC
(num(T)=519; num(N)=44)
JAG1
e
Expression – log2(TPM+1)
HNSC
(num(T)=519; num(N)=44)
HIF1∝
f
Expression – log2(TPM+1)
HNSC
(num(T)=519; num(N)=44)
GJA1
g
Expression – log2(TPM+1)
HNSC
(num(T)=519; num(N)=44)
EST1
h
Expression – log2(TPM+1)
HNSC
(num(T)=519; num(N)=44)
FLT1


a
CD34
HR = 0.56 (0.42 − 0.75)
logrank P = 6.2e−05
Probability
Expression
low
high
Time (months)
Number at risk
low 125 11 1 1 0
high 374 75 15 6 1
b
STAT1
HR = 0.74 (0.56 − 0.97)
logrank P = 0.029
Probability
Expression
low
high
Time (months)
Number at risk
low 296 41 4 1 0
high 203 45 12 6 1
c
CD44
HR = 1.43 (1.07 − 1.92)
logrank P = 0.015
Probability
Expression
low
high
Time (months)
Number at risk
low 192 33 6 2 0
high 307 53 10 5 1
d
EGFR
HR = 1.42 (1.07 − 1.88)
logrank P = 0.014
Probability
Expression
low
high
Time (months)
Number at risk
low 205 39 7 4 0
high 294 47 9 3 1
e
GJA1
HR = 1.44 (1.02 − 2.03)
logrank P = 0.039
Probability
Expression
low
high
Time (months)
Number at risk
low 124 19 3 1 0
high 375 67 13 6 1
f
HNF4A
HR = 1.43 (1.07 − 1.89)
logrank P = 0.014
Probability
Expression
low
high
Time (months)
Number at risk
low 195 34 6 4 0
high 304 52 10 3 1
g
CCND1
HR = 1.73 (1.28 − 2.35)
logrank P = 0.00034
Probability
Expression
low
high
Time (months)
Number at risk
low 172 31 9 5 1
high 327 55 7 2 0


Tyrosine Receptors Family
EGFR
PDGFRB
CD44
Flt1/VEGFRA
JAG1
Hypoxia
Cell 1
Cell 2
PD1
PDL1
Wnt
NOTCH1-4
Cytoplasm
RAF
PI3K
ROS
MEK
B-catenin
AKT
HIF1a
Regulation of Gap junctions and Intracellular communication
SOX2
ERK1/2
NANOG
OCT4
MTOR
LOX
STAT1
STAT3
NFkB
TCF4
PDL1
STAT1
ERK1 /2
B-catenin
HNF4
LOX
HIF1a
Nucleus
STAT3
SOX2
E-cadherin
ETS1
NANOG
OCT4
Proliferation Survival Angiogenesis
NFkB
Snug
EMT/Metastasis Migration Invasion
Apoptosis
Cancer stem cells
Snail
NFkB
CPB
CREB
GJA1
PDL1
CCND1
Cell Cycle Proliferation Angiogenesis