Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Integrative miRNA Profiling Reveals Molecular Heterogeneity in Head and Neck Squamous Cell Carcinoma.

Antra .

JNU: Jawaharlal Nehru University

Vibha Tandon

vibhadelhi6@gmail.com

JNU: Jawaharlal Nehru University https://orcid.org/0000-0001-6146-937X

Research Article

Keywords: Oral squamous cell carcinoma, Head and neck squamous cell carcinoma, microRNA, miRNA, TCGA-HNSC, hsa-miR-301b, hsa-miR-1277, hsa-miR-519a-1, ROC analysis, Kaplan–Meier survival analysis, Pearson correlation, STRING protein–protein interaction network, Gene Ontology, KEGG pathway enrichment, qRT-PCR, molecular heterogeneity

Posted Date: June 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9929728/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Version of Record: A version of this preprint was published at Applied Biochemistry and Biotechnology on October 26th, 2026. See the published version at https://doi.org/10.1007/s12010-026-05904-6.

Title: Integrative miRNA Profiling Reveals Molecular Heterogeneity in Head and Neck Squamous Cell Carcinoma

Antra.[1], Vibha Tandon[1, 2]

[1]Special Centre for Molecular Medicine, Jawaharlal Nehru University, New Delhi-110067, India.

[2]CSIR-Indian Institute of Chemical Biology, Kolkata, West Bengal-700032, India.

Corresponding author:

Prof. Vibha Tandon, Director, CSIR – Indian Institute of Chemical Biology, 4, Raja S.C. Mullick Road, Jadavpur, Kolkata – 700032, India.

Ph.: Off.: +91-33-2473-5368, +91-33-2413-1157 (O), Fax: +91-33-2473-5197,

Email: director@iicb.res.in, vibhadelhi6@gmail.com, vtandon@mail.jnu.ac.in

Data availability: Supplementary data for the paper is attached in the Supplementary Files. All other raw data are available upon request from the corresponding author.

Abstract:

Background: Head and neck squamous cell carcinoma (HNSCC) is a highly heterogeneous malignancy characterized by extensive molecular dysregulation, including aberrant microRNA (miRNA) expression. Identification of clinically relevant miRNAs and their associated regulatory networks may improve understanding of oral squamous cell carcinoma (OSCC) biology and facilitate biomarker discovery.

Objective: The present study aimed to identify novel dysregulated miRNAs associated with OSCC and to investigate their potential diagnostic relevance, predicted target interactions, and functional pathway associations through integrative transcriptomic and clinical validation analyses.

Methods: Mature miRNA expression profiles from The Cancer Genome Atlas (TCGA)-HNSC cohort were retrieved through the UCSC Xena Browser and analyzed using DESeq2 following anatomical site-based filtering to obtain HPV-negative OSCC samples. Differentially expressed miRNAs were identified using adjusted p-value <0.05 and $|\log2FoldChange| \geq 0.5$ thresholds. Based on differential expression patterns and limited prior characterization in OSCC, hsa-miR-1277, hsa-miR-301b, and hsa-miR-519a-1 were selected for downstream analyses. Kaplan–Meier survival, receiver operating characteristic (ROC), integrative target prediction, Pearson correlation, STRING protein–protein interaction network, Gene Ontology (GO), and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were subsequently performed. Independent validation was carried out using four Gene Expression Omnibus (GEO) cohorts and qRT-PCR analysis in Indian OSCC clinical specimens.

Results: Differential expression analysis identified 1,330 significantly dysregulated miRNAs in OSCC, including 946 upregulated and 384 downregulated transcripts. ROC analysis demonstrated strong discriminatory performance for hsa-miR-301b (AUC = 0.868) and moderate performance for hsa-miR-1277 (AUC = 0.735), whereas hsa-miR-519a-1 exhibited comparatively limited diagnostic utility. Integrative target prediction identified five high-confidence targets each for hsa-miR-1277 and hsa-miR-301b, while hsa-miR-519a-1 demonstrated a broader overlap profile comprising 24 candidate genes. Functional enrichment analyses demonstrated enrichment of neuronal signaling, synaptic regulation, ion channel activity, neuroactive ligand–receptor interaction, and cadherin-associated pathways among candidate target genes. Validation across independent GEO cohorts and clinical specimens demonstrated heterogeneous expression patterns with partial concordance to TCGA-derived observations.

Conclusion: This integrative multi-cohort analysis identifies hsa-miR-1277, hsa-miR-301b, and hsa-miR-519a-1 as novel miRNA candidates associated with OSCC-related molecular dysregulation. The findings further highlight the substantial biological heterogeneity underlying HNSCC and emphasize the challenges associated with translating large-scale transcriptomic signatures into reproducible clinical biomarkers.

Keywords: Oral squamous cell carcinoma; Head and neck squamous cell carcinoma; microRNA; miRNA; TCGA-HNSC; hsa-miR-301b; hsa-miR-1277; hsa-miR-519a-1; ROC analysis; Kaplan–Meier survival analysis; Pearson correlation; STRING protein–protein interaction network; Gene Ontology; KEGG pathway enrichment; qRT-PCR; molecular heterogeneity.

Introduction:

Head and neck squamous cell carcinoma (HNSCC) is one of the most prevalent and biologically aggressive malignancies worldwide, accounting for substantial cancer-associated morbidity and mortality. Despite considerable advances in surgical techniques, radiotherapy, chemotherapy, and targeted therapeutic strategies, the clinical outcome of HNSCC patients remains unsatisfactory, particularly in advanced-stage disease **[1, 2]**. High rates of locoregional recurrence, distant metastasis, therapeutic resistance, and disease relapse continue to pose significant clinical challenges. The aggressive behavior of HNSCC is driven by extensive molecular heterogeneity involving complex genetic, epigenetic, and transcriptional alterations that collectively contribute to tumor initiation, progression, and therapeutic failure **[1, 2]**.

Among the diverse classes of regulatory molecules implicated in cancer biology, microRNAs (miRNAs) have emerged as critical post-transcriptional modulators of gene expression. miRNAs are small endogenous non-coding RNAs, approximately 18–24 nucleotides in length, that regulate target gene expression through mRNA degradation or translational repression **[3, 4]**. Increasing evidence has demonstrated that miRNAs participate in the regulation of multiple cellular processes, including proliferation, apoptosis, differentiation, epithelial–mesenchymal transition (EMT), angiogenesis, invasion, metastasis, and stemness. Dysregulation of miRNA expression has been widely associated with carcinogenesis and tumor progression across multiple human cancers, including HNSCC, where miRNAs may function either as oncogenic miRNAs or tumor suppressors depending upon their target networks and cellular context.

Recent advances in high-throughput sequencing technologies and large-scale cancer genomics initiatives have substantially improved our understanding of miRNA-mediated regulatory mechanisms in HNSCC. Numerous studies have identified aberrantly expressed miRNAs associated with tumor progression, metastasis, recurrence, and patient survival, highlighting their potential utility as diagnostic, prognostic, and therapeutic biomarkers **[5-7]**. However,

despite extensive investigation, the reproducibility and clinical translation of miRNA signatures in HNSCC remain limited. Considerable discrepancies frequently exist among independent studies owing to variations in patient cohorts, tumor subsites, ethnic background, sequencing platforms, analytical methodologies, and intrinsic tumor heterogeneity. Moreover, several differentially expressed miRNAs identified through transcriptomic profiling remain poorly characterized, particularly those with limited functional and clinical evidence in HNSCC **[8]**.

Publicly available transcriptomic repositories such as the The Cancer Genome Atlas and Gene Expression Omnibus provide valuable resources for systematic identification of novel molecular biomarkers in cancer. Integrative bioinformatics approaches combining differential expression profiling, survival analysis, diagnostic evaluation, target prediction, functional enrichment analysis, and protein–protein interaction network analysis have emerged as powerful strategies for elucidating biologically relevant miRNA-associated regulatory networks. Furthermore, validation of computational findings in independent datasets and clinical cohorts is essential to establish biological consistency and translational relevance.

In the present study, we performed an integrative transcriptomic analysis of miRNA expression profiles from the TCGA-HNSC cohort to identify novel dysregulated miRNAs associated with HNSCC. Candidate miRNAs were subsequently evaluated for their potential clinical significance using survival and diagnostic analyses, followed by comprehensive target prediction and functional enrichment analyses to investigate their underlying biological relevance. Additionally, independent validation was carried out using multiple GEO datasets and patient-derived clinical samples to assess expression variability and molecular heterogeneity associated with HNSCC.

Material and Methods:

I. **TCGA-HNSC miRNA Expression Dataset Acquisition and Cohort Stratification:** miRNA expression profiles for head and neck squamous cell carcinoma (HNSC) were retrieved from the UCSC Xena Browser, which hosts harmonized datasets generated by The Cancer Genome Atlas. The dataset entitled “TCGA-HNSC miRNA expression (stem-loop quantification)” comprising 569 samples was downloaded from the Genomic Data Commons (GDC) hub available through GDC Xena Hubs **[9-10]**. Expression values were provided as log2(RPM + 1)-normalized counts generated using Illumina sequencing platforms and were directly utilized

for downstream analyses. Corresponding clinical annotations, including patient identifiers, sample type, survival information, and anatomical site details, were obtained from the Genomic Data Commons Portal **[11]**. Since the TCGA-HNSC cohort includes tumors originating from multiple anatomical subsites within the head and neck region, a stringent filtering strategy was applied to establish an oral squamous cell carcinoma (OSCC)-specific cohort. Samples corresponding to oral cavity subsites, including tongue, buccal mucosa, floor of mouth, alveolar ridge, gingiva, and hard palate, were retained for subsequent analyses, whereas tumors derived from non-oral regions such as the larynx, hypopharynx, and oropharynx were excluded. This filtering approach minimized anatomical heterogeneity and ensured biological relevance for OSCC-focused downstream analyses.

II. **Differential Expression and Transcriptomic Analysis:** Raw mRNA and miRNA count matrices together with associated clinical metadata were imported into R (version 4.2.3) for downstream transcriptomic analyses. Samples with incomplete annotations or extremely low library sizes were excluded following quality assessment. Low-abundance genes and miRNAs with fewer than 10 total counts across all samples were removed prior to normalization and statistical modelling. Differential expression analysis was performed using the DESeq2 framework. Count normalization was conducted using the median-of-ratios method implemented in DESeq2 to account for sequencing depth variability across samples. Gene-wise dispersion estimates were calculated using empirical Bayes shrinkage, and differential expression testing was performed using negative binomial generalized linear models with tumor versus normal tissue status specified as the primary experimental condition. For downstream visualization and comparative analyses, normalized count matrices were transformed using variance stabilizing transformation (VST) or regularized log (rlog) transformation to reduce heteroscedasticity associated with count-based sequencing data. Normalized expression values were subsequently utilized for clustering, correlation, survival, and receiver operating characteristic (ROC) analyses.

III. **Kaplan–Meier Survival Analysis:** Kaplan–Meier survival analysis was performed to evaluate the prognostic significance of candidate miRNAs identified through TCGA-based OSCC transcriptomic profiling. Normalized miRNA expression data and corresponding clinical annotations for the TCGA-HNSC cohort were obtained from the UCSC Xena Browser. Expression profiles corresponding to hsa-miR-1277, hsa-miR-519a-1, and hsa-miR-301b were extracted for downstream analyses. Patient identifiers derived from TCGA sample barcodes

were matched with corresponding clinical metadata obtained from the Genomic Data Commons Portal. Overall survival (OS) was defined as the interval between diagnosis and death or last follow-up. Patients without documented death events were considered censored. For each miRNA, patients were stratified into high- and low-expression groups based on the median expression value across the cohort. Kaplan–Meier survival curves were generated using the survival and survminer packages in R (version 4.4.1), and differences in survival distributions were assessed using the log-rank test. Hazard ratios (HRs) together with 95% confidence intervals (CIs) were estimated using univariate Cox proportional hazards regression analysis. A two-tailed p value < 0.05 was considered statistically significant **[11-14]**.

IV. **Receiver Operating Characteristic (ROC) Curve Analysis:** The diagnostic performance of candidate miRNAs was evaluated using receiver operating characteristic (ROC) curve analysis in the TCGA-HNSC cohort. Mature miRNA expression profiles corresponding to hsa-miR-1277, hsa-miR-301b, and hsa-miR-519a-1 were retrieved from normalized TCGA miRNA datasets available through the UCSC Xena Browser. Samples were categorized according to TCGA barcode annotations, where "01A" represented primary tumor tissues and "11A" represented adjacent normal tissues. Following exclusion of incomplete observations, a total of 562 samples comprising 518 tumor and 44 normal tissues were included for downstream analysis. ROC analysis was performed using the pROC package in R. ROC curves were generated using non-parametric empirical ROC estimation, and diagnostic performance was evaluated by calculating the area under the curve (AUC) together with sensitivity and specificity distributions across classification thresholds **[15]**. Statistical significance was determined using two-tailed testing, and graphical visualization of ROC curves was performed using ggplot2-based plotting functions.

V. **GEO Microarray Validation Cohorts:** To validate the reproducibility of differentially expressed miRNAs identified from TCGA miRNA-sequencing data, four independent microarray-based datasets, including GSE45238, GSE28100, GSE31277, and GSE124566, were retrieved from the Gene Expression Omnibus (GEO) database and analyzed as external validation cohorts **[16]**. These datasets were selected to evaluate consistency of candidate miRNA expression patterns across independent patient cohorts and experimental platforms. Differential expression analysis for each dataset was performed using the GEO2R platform implementing the limma framework with empirical Bayes moderation. Statistical significance was determined using Benjamini–Hochberg false discovery rate (FDR) correction **[17-19]**.

Differentially expressed miRNAs were identified using an adjusted p-value threshold < 0.05 together with a |log2 fold change| cutoff ≥ 0.25. Volcano plots were generated to visualize the distribution of significantly upregulated and downregulated miRNAs. Candidate miRNAs identified from TCGA analyses were subsequently cross-validated across all GEO datasets to assess reproducibility and directional consistency of expression patterns.

VI. **miRNA Target Prediction and Pearson Correlation Analysis:** Candidate target genes associated with hsa-miR-1277, hsa-miR-519a-1, and hsa-miR-301b were identified through integrative analysis incorporating differential gene expression profiles together with multiple in silico miRNA target prediction databases, including TargetScan, miRDB, and miRWalk. Predicted targets were intersected with significantly downregulated genes identified from TCGA-HNSC transcriptomic datasets to prioritize biologically relevant candidate targets **[20-22]**. Pearson correlation analysis was subsequently performed to evaluate the relationship between candidate miRNA expression and corresponding target gene expression within the TCGA-HNSC cohort. Normalized expression matrices for candidate miRNAs and target genes were extracted and analyzed in R using the ggplot2 framework **[17]**. Correlation coefficients (Pearson's r) together with associated p values were calculated using two-tailed Pearson correlation testing. Scatter plots were generated with linear regression fitting and 95% confidence intervals to visualize inverse or direct associations between miRNA and target gene expression levels.

VII. **STRING Protein–Protein Interaction Network Analysis:** To investigate the functional interaction landscape of candidate miRNA target genes, protein–protein interaction (PPI) network analysis was performed using the STRINGdb package in R. Candidate target genes identified through integrative transcriptomic analysis, target prediction databases, and Pearson correlation analysis were included for network generation. STRING database version 12.0 was queried using Homo sapiens (taxonomy ID: 9606) as the reference organism with a medium confidence interaction score threshold of 0.4. Genes were mapped to STRING identifiers, and interaction pairs among queried genes were extracted to construct the interaction network **[23, 24]**. The resulting PPI network was visualized using the igraph package in R employing the Fruchterman–Reingold layout algorithm to identify interaction connectivity and functional clustering patterns among candidate target genes.

VIII. **Gene Ontology (GO) and KEGG Pathway Enrichment Analysis:** To investigate the biological functions and signaling pathways associated with candidate miRNA target genes, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were performed using the clusterProfiler package in R. Target genes identified through integrative transcriptomic analysis and Pearson correlation analysis for hsa-miR-1277, hsa-miR-301b, and hsa-miR-519a-1 were used as input gene sets. Gene symbols were converted into Entrez Gene IDs using the org.Hs.eg.db database. GO enrichment analysis was conducted across Biological Process (BP), Cellular Component (CC), and Molecular Function (MF) ontologies, while KEGG pathway enrichment analysis was performed to identify significantly associated signaling pathways **[25]**. Enrichment significance was evaluated using adjusted p-values, and enriched terms were ranked according to gene count and statistical significance. Enrichment results were visualized as bar plots using ggplot2 and enrichplot packages, where bar length represented gene count and color intensity corresponded to −log10(adjusted p-value).

IX. **Clinical Sample Collection and Quantitative Real-Time PCR Validation:** OSCC tissue specimens together with adjacent normal oral mucosal tissues were collected for experimental validation of candidate miRNA expression. Total RNA, including small RNAs, was isolated using TRIzol Reagent (Thermo Fisher Scientific, USA) according to the manufacturer's instructions. Reverse transcription was performed using the miR-Q assay strategy in combination with the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems, USA), as previously described **[26, 27]**. Quantitative real-time PCR (qRT-PCR) was performed using SYBR Green chemistry on an Applied Biosystems StepOnePlus Real-Time PCR System. Expression levels of hsa-miR-1277, hsa-miR-519a-1, and hsa-miR-301b were normalized against RNU6B, which served as the endogenous internal control. Relative expression levels were calculated using the 2^−ΔΔCt method. All reactions were performed in triplicate, and data are presented as mean ± standard deviation (SD). Statistical analysis was performed using a one-sample t-test against a theoretical mean of zero, and a two-tailed p value < 0.05 was considered statistically significant.

Stemloop mir-519a-1:

GAAAGAAGGCGAGGAGCAGATCGAGGAAGAAGACGGAAGAATGTGCGTCTCGCCTTCTTTCACAC
TCTAA

FP mir-519a-1-3p:
GCAAAGTGCATCCTTTTAGAGTG

Stemloop mir-301b-3p:
GAAAGAAGGCGAGGAGCAGATCGAGGAAGAAGACGGAAGAATGTGCGTCTCGCCTTCTTTCGCTT
TGACA

FP Stem mir-301b-3p:
GCCAGTGCAATGATATTGTCA

Universal RP Stemloop:
CGAGGAAGAAGACGGAAGAAT

Stemloop miR-1277-5p:
GAAAGAAGGCGAGGAGCAGATCGAGGAAGAAGACGGAAGAATGTGCGTCTCGCCTTCTTTCATAC
GTACA

FP Stem miR-1277-5p:
GCGGAAATATATATATATATGTACGTAT

FP Stem U6: GCTTCGGCAGCACATATACTAAAAT

RP Stem U6: CGCTTCACGAATTTGCGTGTCAT

Results:

1. Differential miRNA Expression Profiling Identifies Candidate Oncogenic miRNAs in OSCC:

Mature miRNA expression profiles from the TCGA-HNSC cohort were retrieved through the UCSC Xena Browser and filtered to establish an HPV-negative OSCC-specific dataset comprising 323 samples, including 304 tumor tissues and 19 normal oral mucosa controls **(Supplementary Document 1)**. Differential expression analysis performed using DESeq2 identified 1,330 significantly dysregulated miRNAs in OSCC, including 946 upregulated and 384 downregulated transcripts relative to normal oral tissues (FDR ≤ 0.05; log2FC ≥ 0.5).

Global differential miRNA expression patterns were visualized using a volcano plot **(Figure 1)**. Among the significantly deregulated miRNAs, hsa-miR-1277, hsa-miR-301b, and hsa-miR-519a-1 demonstrated prominent upregulation in OSCC tissues and were prioritized for subsequent diagnostic, prognostic, and functional analyses based on their differential expression profiles and potential biological relevance.

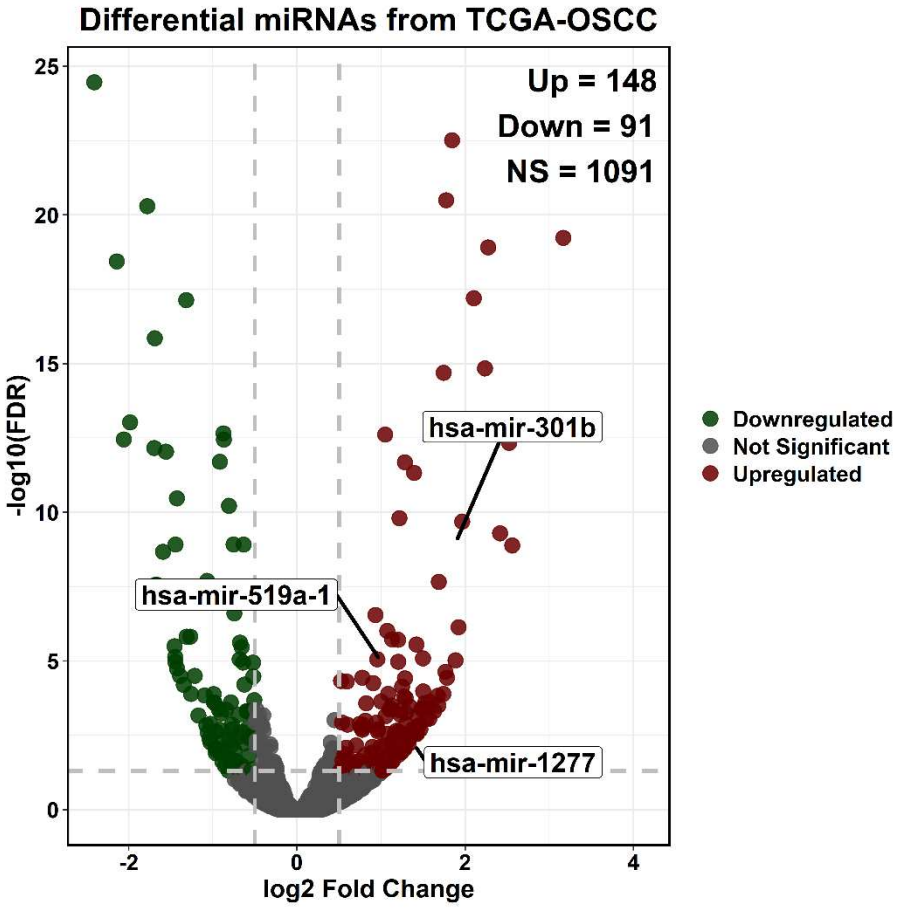


Figure 1. Differential miRNA expression landscape in TCGA-derived OSCC cohort. Volcano plot illustrating differentially expressed mature miRNAs identified in HPV-negative OSCC samples derived from the TCGA-HNSC cohort obtained through the UCSC Xena Browser. Differential expression analysis was performed using DESeq2 with significance thresholds set at FDR ≤ 0.05 and log2 fold change ≥ 0.5. A total of 1,330 significantly dysregulated miRNAs were identified, including 946 upregulated and 384 downregulated transcripts in OSCC tissues relative to normal oral mucosa controls. Candidate oncogenic miRNAs, including hsa-miR-1277, hsa-miR-301b, and hsa-miR-519a-1, are highlighted for downstream validation and functional characterization.

2. Prognostic and Diagnostic Assessment of Candidate miRNAs in the TCGA-HNSC Cohort:

To further investigate the translational relevance of the shortlisted candidate miRNAs identified through differential expression profiling, Kaplan–Meier survival analysis and receiver operating characteristic (ROC) curve analysis were performed using the TCGA-HNSC cohort **(Figure 2A–F)**.

Kaplan–Meier survival analysis was conducted to evaluate the prognostic significance of hsa-miR-1277, hsa-miR-519a-1, and hsa-miR-301b in HNSCC patients **(Figure 2A–C)**. Patients

were stratified into high- and low-expression groups based on median miRNA expression levels **(Supplementary document 2)**. hsa-miR-519a-1 expression was not significantly associated with overall survival (HR = 0.924, 95% CI: 0.708–1.205; log-rank p = 0.559). Similarly, no significant survival association was observed for hsa-miR-1277 (HR = 1.000, 95% CI: 0.767–1.303; log-rank p = 0.998) or hsa-miR-301b (HR = 1.039, 95% CI: 0.798–1.354; log-rank p = 0.774). These findings indicate that although the candidate miRNAs were significantly dysregulated in OSCC tissues, their expression levels did not demonstrate prognostic significance for overall survival within the heterogeneous TCGA-HNSC cohort.

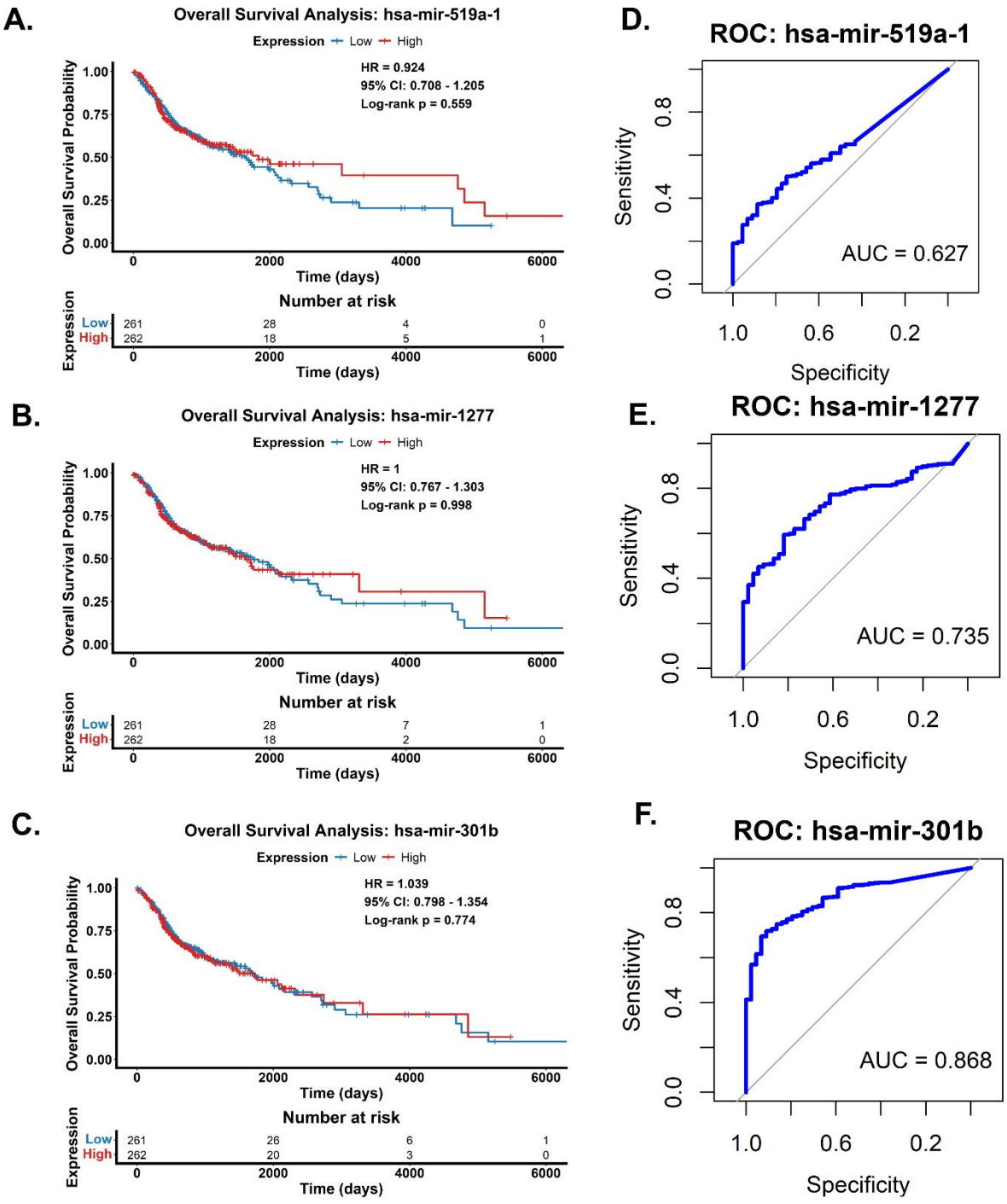


Figure 2. Survival and ROC curve analyses of candidate miRNAs in the TCGA-HNSC cohort. (A–C) Kaplan–Meier survival curves illustrating overall survival stratified according to high- and low-expression groups for (A) hsa-miR-1277, (B) hsa-miR-519a-1, and (C) hsa-miR-301b in the TCGA-HNSC cohort. Patients were dichotomized based on median miRNA expression values. Hazard ratios (HRs), 95% confidence intervals (CIs), and log-rank p values derived from Cox proportional hazards regression analysis are indicated within each panel. None of the evaluated candidate miRNAs demonstrated statistically significant prognostic association with overall survival. (D–F) Receiver operating characteristic (ROC) curve analysis evaluating the discriminatory performance of candidate miRNAs using normalized TCGA-HNSC miRNA expression profiles comprising 518 tumor and 44

normal tissue samples. (D) hsa-miR-519a-1 demonstrated comparatively lower discriminatory performance with an area under the curve (AUC) of 0.627. (E) hsa-miR-1277 exhibited moderate diagnostic accuracy with an AUC of 0.735. (F) hsa-miR-301b showed strong discriminatory performance with an AUC of 0.868. Collectively, these findings indicate superior diagnostic potential of hsa-miR-301b, followed by hsa-miR-1277, whereas hsa-miR-519a-1 displayed comparatively weaker discriminatory utility in the TCGA-HNSC cohort.

Given the limited prognostic association, the diagnostic utility of these candidate miRNAs was subsequently assessed using ROC curve analysis based on normalized miRNA expression profiles derived from 518 tumor and 44 normal tissue samples **(Figure 2D–F)**. Among the evaluated candidates, hsa-miR-301b demonstrated the strongest discriminatory performance with an area under the curve (AUC) of 0.868, indicating high diagnostic accuracy for distinguishing HNSCC tissues from normal controls. hsa-miR-1277 also exhibited appreciable discriminatory capability with an AUC of 0.735, suggesting moderate diagnostic performance **(Supplementary Document 3)**. In contrast, hsa-miR-519a-1 demonstrated comparatively lower diagnostic accuracy with an AUC of 0.627, indicating limited standalone discriminatory utility. Collectively, these findings support the potential diagnostic relevance of hsa-miR-301b and hsa-miR-1277 in HNSCC and further validate the transcriptomic alterations identified through differential expression profiling.

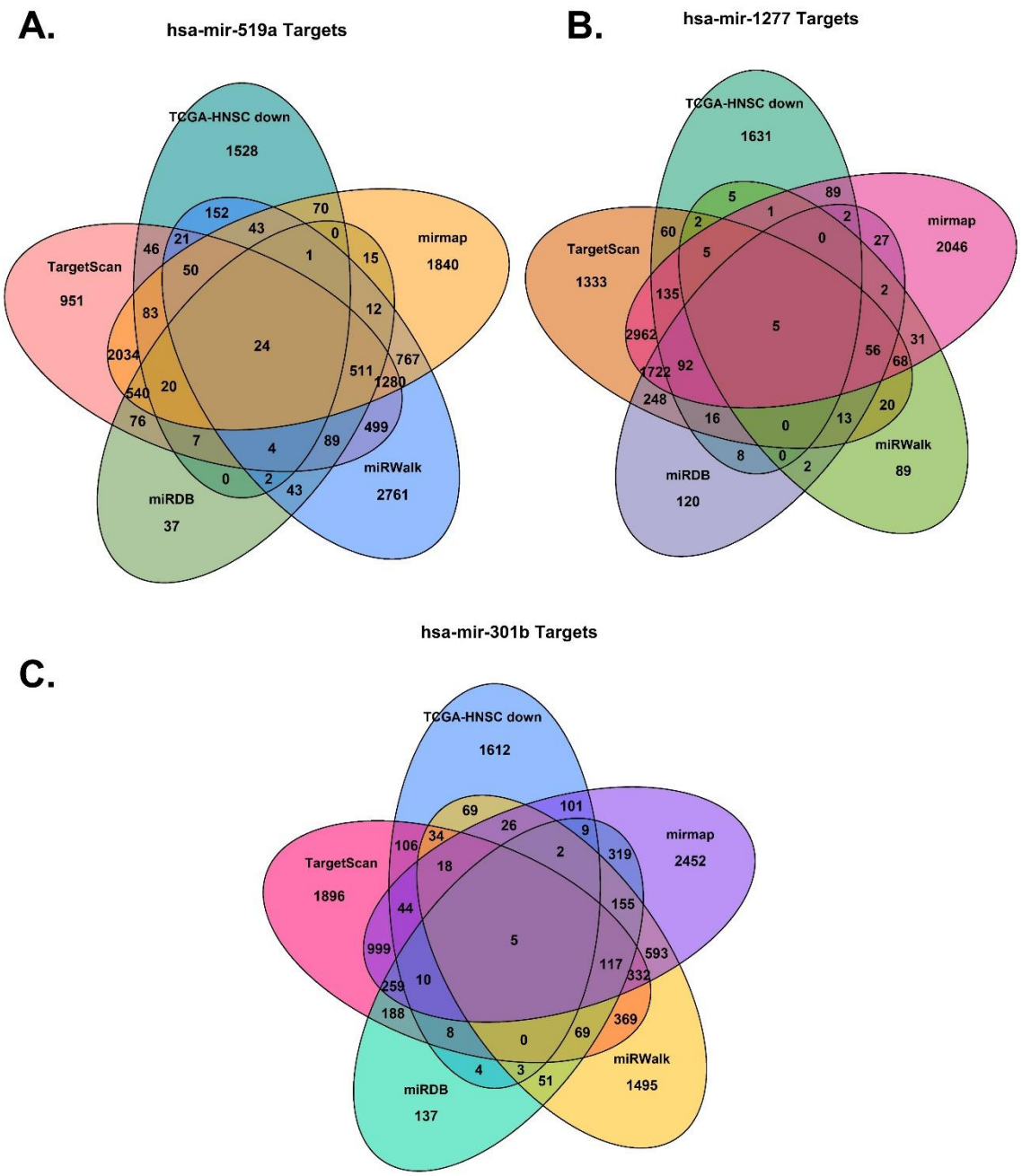


Figure 3. Integrative overlap analysis of predicted target genes associated with candidate miRNAs in OSCC. Venn diagram analyses illustrating the overlap between significantly downregulated genes identified from the TCGA-HNSC dataset and predicted miRNA targets obtained from TargetScan, miRDB, miRWalk, and miRmap databases. (A) Overlap profile of hsa-miR-519a identifying 24 common candidate target genes shared across all datasets. (B) Overlap analysis of hsa-miR-1277 identifying five common candidate target genes, including KCNJ3, DLG2, KLHL14, AGTR1, and HLF. (C) Overlap analysis of hsa-miR-301b identifying five common candidate target genes comprising DLG2, GABRB2, NR3C2, PRKAA2, and AR. Intersecting genes represent high-confidence putative downstream targets consistently identified across transcriptomic and in silico prediction datasets.

3. **Identification of High-Confidence Candidate Targets of Dysregulated miRNAs in OSCC:**

To further investigate the potential downstream regulatory landscape associated with the shortlisted candidate miRNAs, an integrative target prediction analysis was performed for hsa-miR-1277, hsa-miR-519a, and hsa-miR-301b using TCGA-HNSC transcriptomic data in combination with multiple independent in silico target prediction databases **(Figure 3A–C)**. Differentially downregulated genes identified from the TCGA-HNSC dataset were intersected with predicted miRNA targets obtained from TargetScan, miRDB, miRWalk, and miRmap databases to prioritize biologically relevant candidate targets **(Supplementary Document 4-6)**. Venn diagram analyses were generated in R using the VennDiagram package, while data preprocessing and import were performed using the readxl package.

Integrative overlap analysis for hsa-miR-1277 identified five common candidate target genes, namely *KCNJ3, DLG2, KLHL14, AGTR1,* and *HLF,* shared across TCGA-HNSC downregulated genes and all four prediction databases **(Figure 3B)**. These overlapping genes represented high-confidence putative downstream targets of hsa-miR-1277 and were prioritized for subsequent correlation and functional analyses.

Similarly, overlap analysis for hsa-miR-301b identified five common candidate target genes, including *DLG2, GABRB2, NR3C2, PRKAA2,* and *AR* **(Figure 3C)**. Several of these genes are associated with signal transduction, transcriptional regulation, metabolic homeostasis, and cellular signaling pathways, suggesting a potential regulatory role of hsa-miR-301b in OSCC-associated oncogenic processes.

In contrast, hsa-miR-519a demonstrated a substantially broader overlap profile, with 24 common candidate target genes identified across TCGA-HNSC downregulated genes and all four prediction databases **(Figure 3A)**. These included *PCDH20, CA10, PCDHA12, PCDHA13, TOX3, FBXO40, CHRM2, LMX1A, GRIK3, PCDHAC1, KCNB1, SLITRK3, PPP1R9A, PCDHA11, LONRF2, FAM216B, SNX31, PCDHA3, HFM1, LRP1B, PCDHA4, GPM6A, RBM20, and PCDHAC2*. Notably, several identified targets belonged to the protocadherin family and neuronal signaling-associated pathways, suggesting that hsa-miR-519a may regulate diverse cellular processes linked to adhesion dynamics, intercellular signaling, and tumor progression in OSCC.

Collectively, the integrative intersection strategy combining TCGA-HNSC downregulated genes with multiple independent target prediction algorithms enabled the identification of robust and biologically relevant candidate targets for all three miRNAs. These high-confidence target genes were subsequently prioritized for downstream Pearson correlation, pathway enrichment, and functional network analyses.

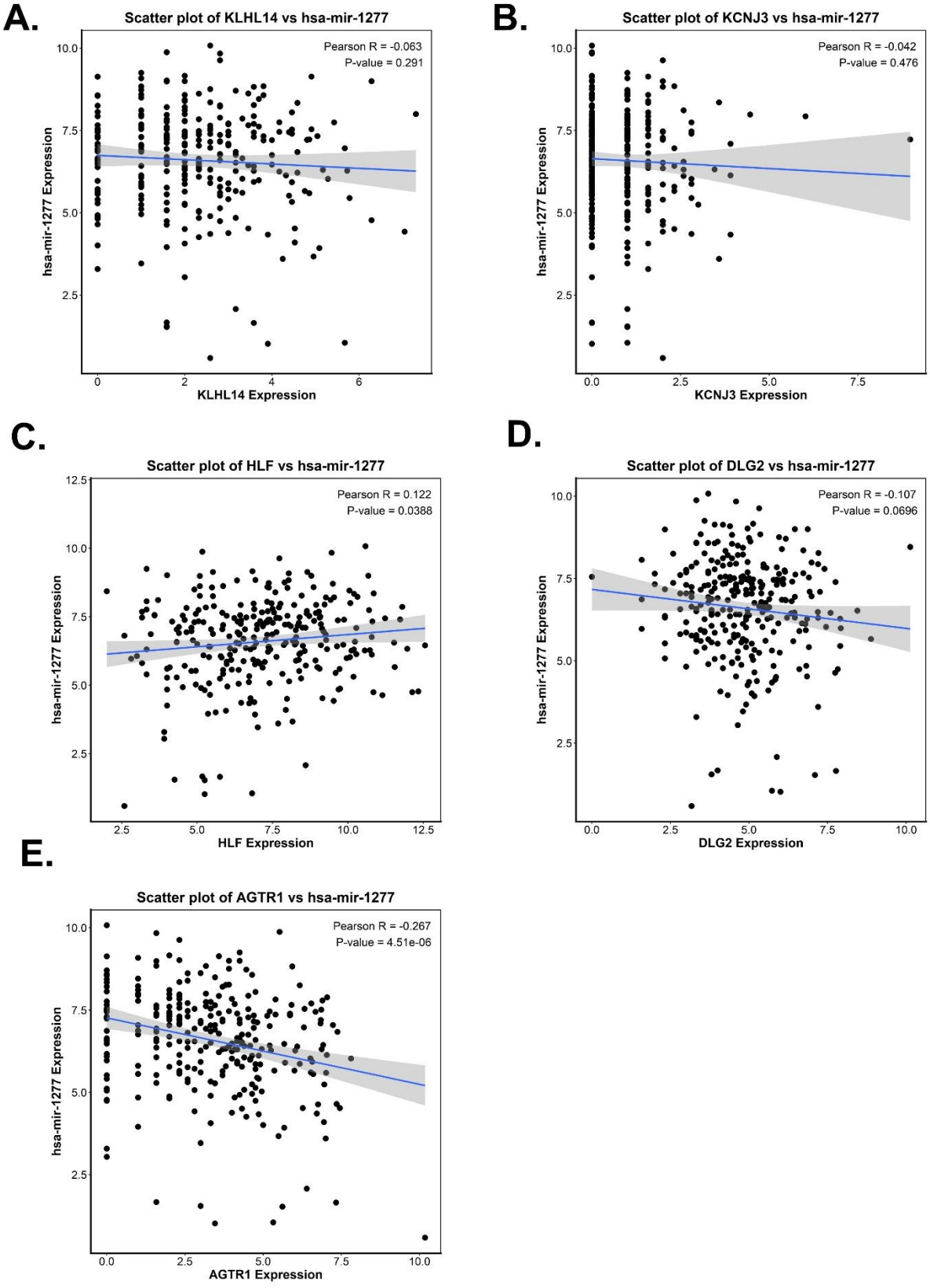


Figure 4. Pearson correlation analysis between hsa-miR-1277 and predicted target genes in OSCC. Scatter plots illustrating Pearson correlation analysis between hsa-miR-1277 expression and the predicted target genes AGTR1, DLG2, HLF, KCNJ3, and KLHL14 in TCGA-HNSC tumor samples. Each dot represents an individual patient sample, and linear regression lines with 95% confidence intervals are shown. hsa-miR-1277 demonstrated

a significant inverse correlation with AGTR1 ($R = -0.267$, $p = 4.51 \times 10^{-6}$), whereas weak negative correlations were observed with DLG2 ($R = -0.1073$, $p = 0.0696$), KCNJ3 ($R = -0.0423$, $p = 0.4758$), and KLHL14 ($R = -0.0626$, $p = 0.2906$). In contrast, HLF exhibited a weak positive correlation with hsa-miR-1277 expression ($R = 0.122$, $p = 0.0388$).

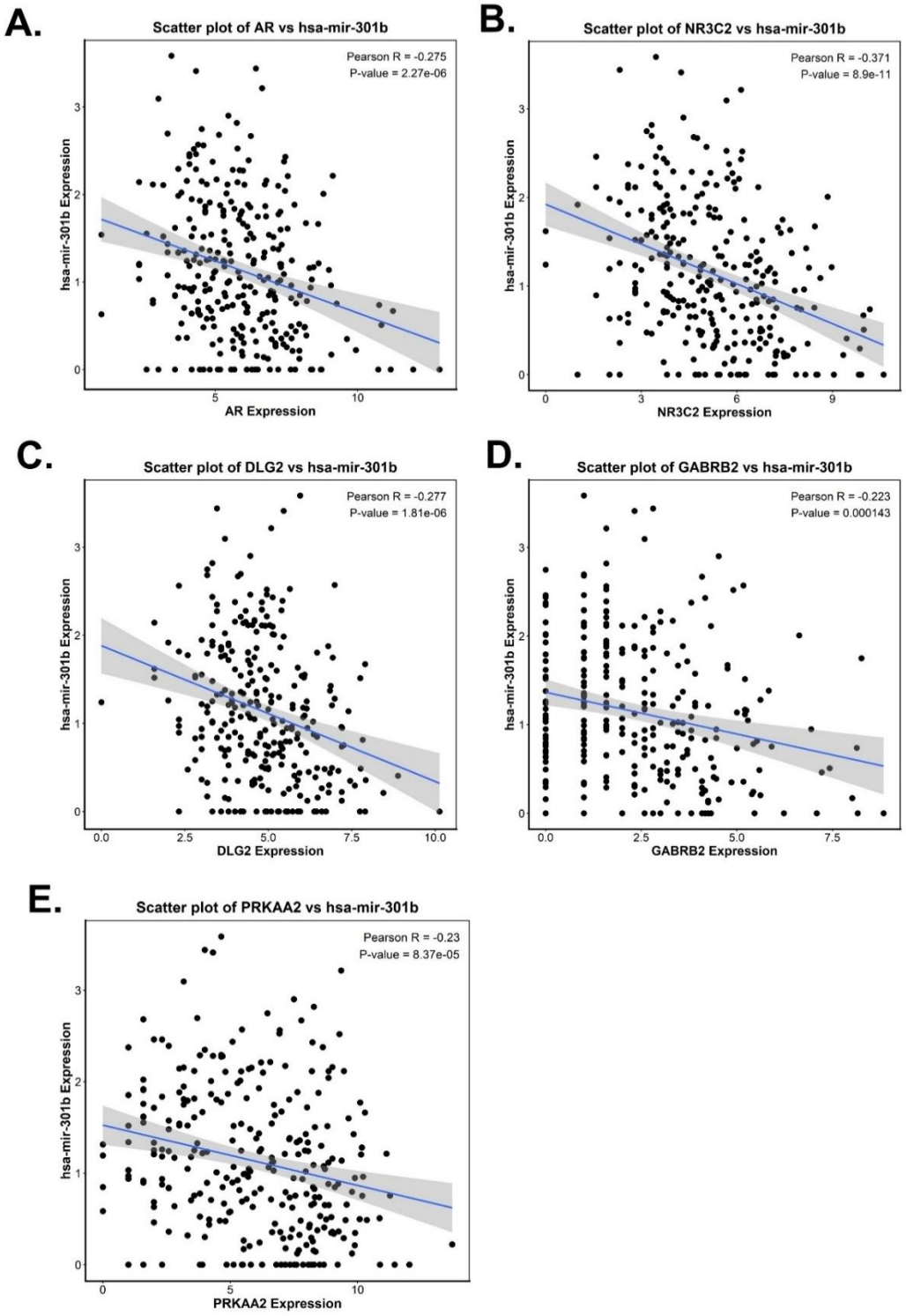


Figure 5. Pearson correlation analysis between hsa-miR-301b and predicted target genes in OSCC. Scatter plots showing Pearson correlation analysis between hsa-miR-301b expression and the predicted target genes AR,

DLG2, GABRB2, NR3C2, and PRKAA2 in TCGA-HNSC tumor samples. Each dot represents an individual tumor sample, and linear regression lines with 95% confidence intervals are displayed. hsa-miR-301b exhibited significant inverse correlations with AR ($R = -0.2749$, $p = 2.27 \times 10^{-6}$), DLG2 ($R = -0.2774$, $p = 1.81 \times 10^{-6}$), GABRB2 ($R = -0.2226$, $p = 1.43 \times 10^{-4}$), NR3C2 ($R = -0.3707$, $p = 8.90 \times 10^{-11}$), and PRKAA2 ($R = -0.2301$, $p = 8.37 \times 10^{-5}$). Among the analyzed targets, NR3C2 demonstrated the strongest inverse association with hsa-miR-301b expression.

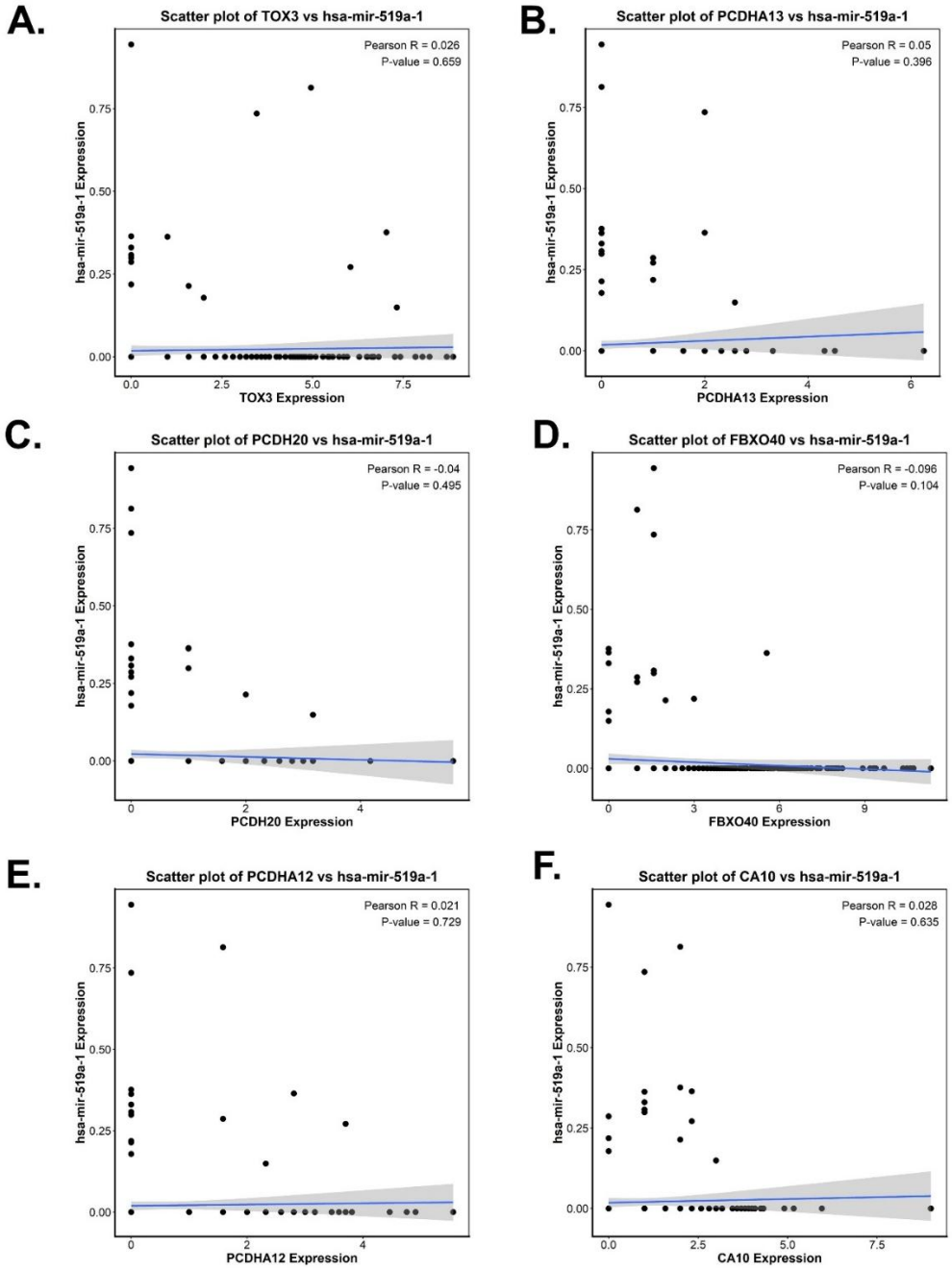


Figure 6. Pearson correlation analysis between hsa-miR-519a-1 and predicted target genes in OSCC. Scatter plots depicting Pearson correlation analysis between hsa-miR-519a-1 expression and the predicted target genes CA10, FBXO40, PCDH20, PCDHA12, PCDHA13, and TOX3 in TCGA-HNSC tumor samples. Each dot

represents an individual patient sample, and linear regression lines with 95% confidence intervals are shown. hsa-miR-519a-1 demonstrated weak correlations with all analyzed target genes, including CA10 ($R = 0.0281$, $p = 0.6353$), FBXO40 ($R = -0.0961$, $p = 0.1043$), PCDH20 ($R = -0.0404$, $p = 0.4951$), PCDHA12 ($R = 0.0206$, $p = 0.7285$), PCDHA13 ($R = 0.0503$, $p = 0.3957$), and TOX3 ($R = 0.0262$, $p = 0.6587$). No statistically significant correlations were identified between hsa-miR-519a-1 expression and the analyzed candidate target genes within the TCGA-HNSC cohort.

4. **Pearson Correlation Analysis Supports Putative miRNA–Target Regulatory Associations in OSCC:** Following the identification of high-confidence candidate target genes through integrative overlap analysis, Pearson correlation analysis was subsequently performed to further investigate the potential regulatory association between the shortlisted miRNAs and their predicted downstream targets within the TCGA-HNSC transcriptomic cohort **(Figures 4–6)**. Since miRNA-mediated gene regulation is frequently associated with transcript destabilization and reduced target mRNA abundance, inverse expression correlations between candidate miRNAs and their predicted targets were considered supportive of potential post-transcriptional regulatory interactions in OSCC **(Supplementary Document 7)**.

   For hsa-miR-1277, AGTR1 expression demonstrated a significant inverse correlation with miRNA levels ($R = -0.267$, $p = 4.51 \times 10^{-6}$), supporting a potential miRNA-mediated suppressive interaction in OSCC **(Figure 4)**. In contrast, DLG2, KCNJ3, and KLHL14 exhibited weak negative correlations that did not reach statistical significance. Interestingly, HLF displayed a weak but statistically significant positive correlation with hsa-miR-1277 expression ($R = 0.122$, $p = 0.0388$), suggesting the possibility of indirect or context-dependent regulatory interactions within the tumor transcriptomic landscape.

   In the case of hsa-miR-301b, all analyzed candidate target genes exhibited significant negative correlations with miRNA expression (Figure 5). Among these, NR3C2 demonstrated the strongest inverse association ($R = -0.3707$, $p = 8.90 \times 10^{-11}$), followed by DLG2 ($R = -0.2774$, $p = 1.81 \times 10^{-6}$) and AR ($R = -0.2749$, $p = 2.27 \times 10^{-6}$). Significant negative correlations were also observed for GABRB2 and PRKAA2. The consistent inverse correlation pattern across multiple predicted targets strongly supports a potential oncogenic role of hsa-miR-301b through coordinated repression of tumor suppressive and regulatory genes in OSCC.

   In contrast, hsa-miR-519a-1 did not demonstrate significant transcript-level correlations with the analyzed candidate targets, including CA10, FBXO40, PCDH20, PCDHA12, PCDHA13, and TOX3 (Figure 6). Although weak positive and negative trends were observed, none

achieved statistical significance. These findings suggest that the predicted targets of hsa-miR-519a-1 may be regulated through alternative mechanisms, including context-specific post-transcriptional modulation or indirect regulatory interactions not fully reflected at the bulk transcriptomic level within the TCGA-HNSC cohort.

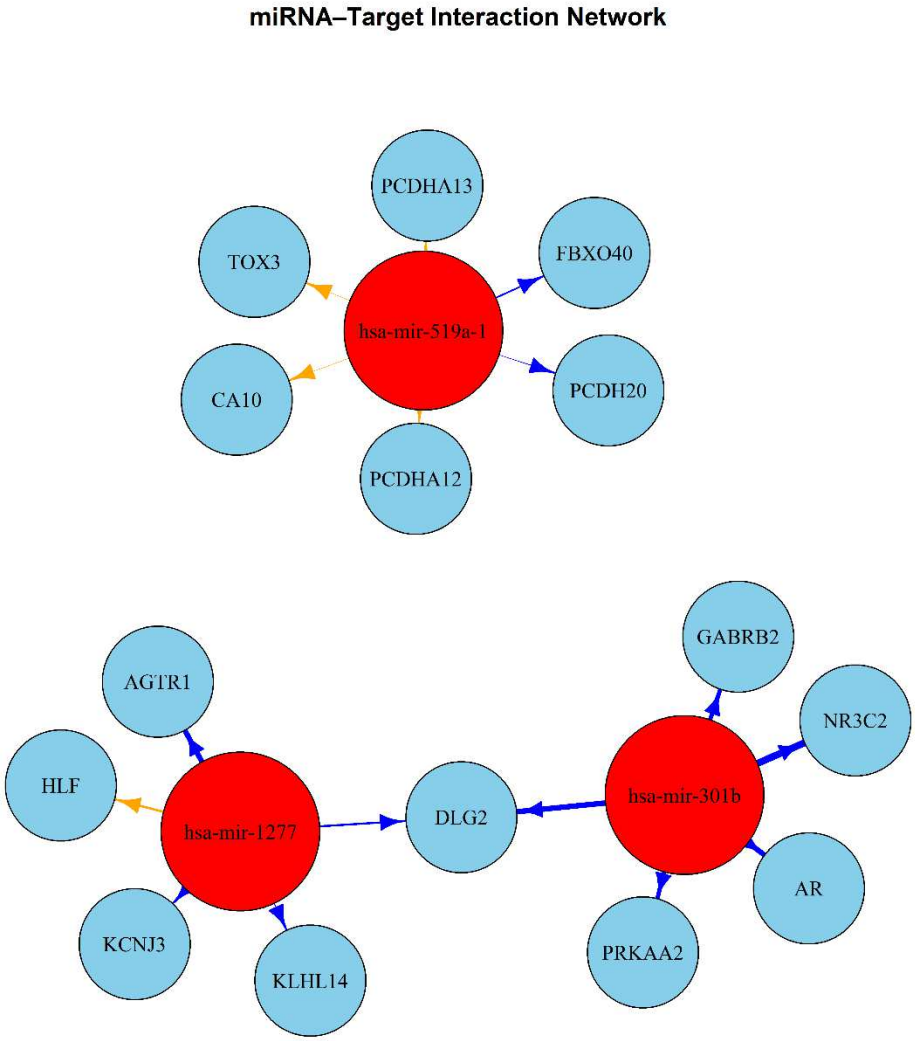


Figure 7. Protein–protein interaction network of candidate target genes associated with hsa-miR-1277, hsa-miR-301b, and hsa-miR-519a-1. STRING database (version 12.0)-based protein–protein interaction (PPI) network illustrating functional associations among candidate target genes identified through integrative transcriptomic and target prediction analyses. Nodes represent individual target genes, while connecting edges indicate predicted or experimentally supported functional interactions. Distinct interaction clusters were observed among targets associated with hsa-miR-301b and protocadherin-associated targets of hsa-miR-519a-1. DLG2 appeared as a shared interacting node between hsa-miR-1277 and hsa-miR-301b target networks, suggesting potential convergence of regulatory pathways. Several isolated or weakly connected nodes were also observed, indicating functional heterogeneity among the predicted miRNA target genes.

5. **Functional Interaction Network Analysis Reveals Connectivity Among Candidate miRNA Target Genes:** To further explore the potential biological relationships among the identified candidate target genes, STRING-based protein–protein interaction (PPI) network analysis was performed using targets associated with hsa-miR-1277, hsa-miR-301b, and hsa-miR-519a-1 **(Figure 7)**. Since miRNA-mediated oncogenic regulation frequently involves coordinated modulation of interconnected signaling networks rather than isolated genes, functional interaction analysis was undertaken to evaluate whether the predicted targets participated in shared regulatory pathways or molecular interaction modules within OSCC-associated transcriptomic networks.

   The resulting interaction network demonstrated moderate but distinct clustering patterns among several candidate targets, whereas multiple genes remained weakly connected or isolated, indicating functional heterogeneity among the predicted miRNA-regulated genes **(Supplementary Document 8)**. Among the analyzed targets, genes associated with hsa-miR-301b exhibited comparatively stronger interaction connectivity. DLG2 emerged as a shared interacting node linking targets associated with both hsa-miR-1277 and hsa-miR-301b, suggesting potential convergence of these miRNAs on common regulatory pathways. Additional interactions involving NR3C2, AR, PRKAA2, and GABRB2 further supported the involvement of signaling pathways associated with cellular regulation, signal transduction, and transcriptional control.

   The hsa-miR-519a-1-associated target network demonstrated partial clustering predominantly among protocadherin family members, including PCDHA12 and PCDHA13, together with neuronal and adhesion-associated targets such as TOX3, FBXO40, and PCDH20. However, several predicted targets displayed limited or absent interactions within the STRING database, resulting in a comparatively sparse interaction architecture. Collectively, the STRING-based interaction analysis suggested potential involvement of cell adhesion-associated processes, neuronal signaling pathways, and regulatory signaling networks in OSCC-associated miRNA dysregulation.

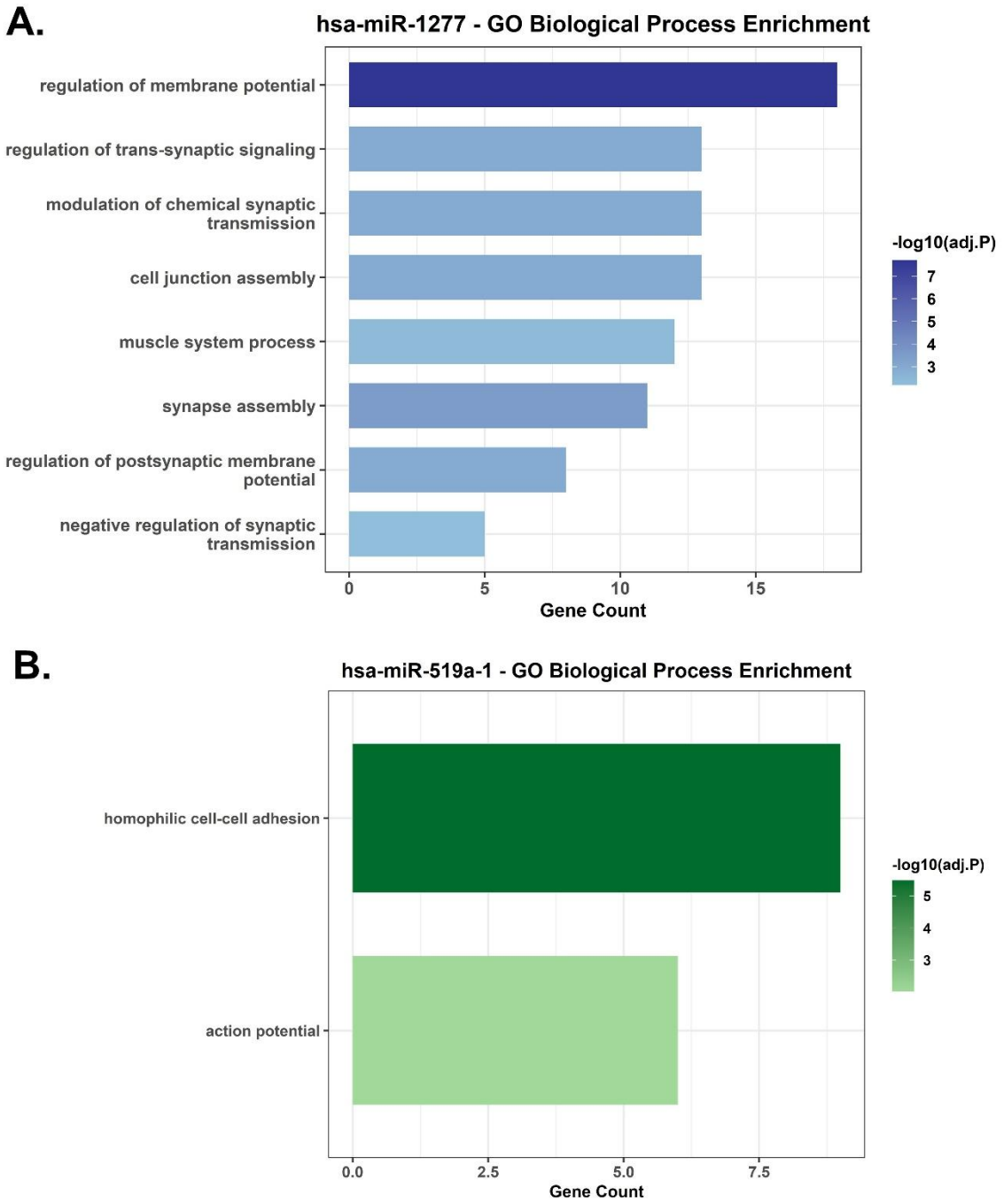


Figure 8. GO Biological Process enrichment analysis of candidate miRNA target genes. (A) Gene Ontology (GO) Biological Process enrichment analysis of hsa-miR-1277-associated target genes demonstrating significant enrichment of pathways related to regulation of membrane potential, trans-synaptic signaling, synapse assembly, and modulation of chemical synaptic transmission. **(B)** GO Biological Process enrichment analysis of hsa-miR-519a-1-associated target genes showing enrichment of homophilic cell–cell adhesion and action potential-associated biological processes. Bar length represents the number of genes associated with each enriched term, while color intensity corresponds to enrichment significance represented as −log10 (adjusted p-value).

6. **Functional Enrichment Analysis Reveals Synaptic and Cell Adhesion-Associated Pathways Linked to Candidate miRNA Targets:** To further investigate the biological

processes and signaling pathways potentially regulated by the identified candidate miRNAs, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were performed using the selected target genes derived from integrative TCGA-HNSC transcriptomic and in silico target prediction analyses **(Figures 8–9)**. Given the moderate interaction clustering observed in STRING network analysis, enrichment analysis was subsequently undertaken to determine whether the identified target genes converged on specific biological pathways relevant to OSCC pathobiology.

GO Biological Process enrichment analysis of hsa-miR-1277-associated target genes demonstrated significant enrichment of neuronal and synaptic regulatory processes, including regulation of membrane potential, regulation of trans-synaptic signaling, modulation of chemical synaptic transmission, synapse assembly, and regulation of postsynaptic membrane potential **(Figure 8A)**. Additional enrichment of cell junction assembly and muscle system-associated processes further suggested potential involvement of hsa-miR-1277-regulated targets in membrane-associated signaling dynamics and intercellular communication mechanisms implicated in tumor progression **(Supplementary Document 9)**.

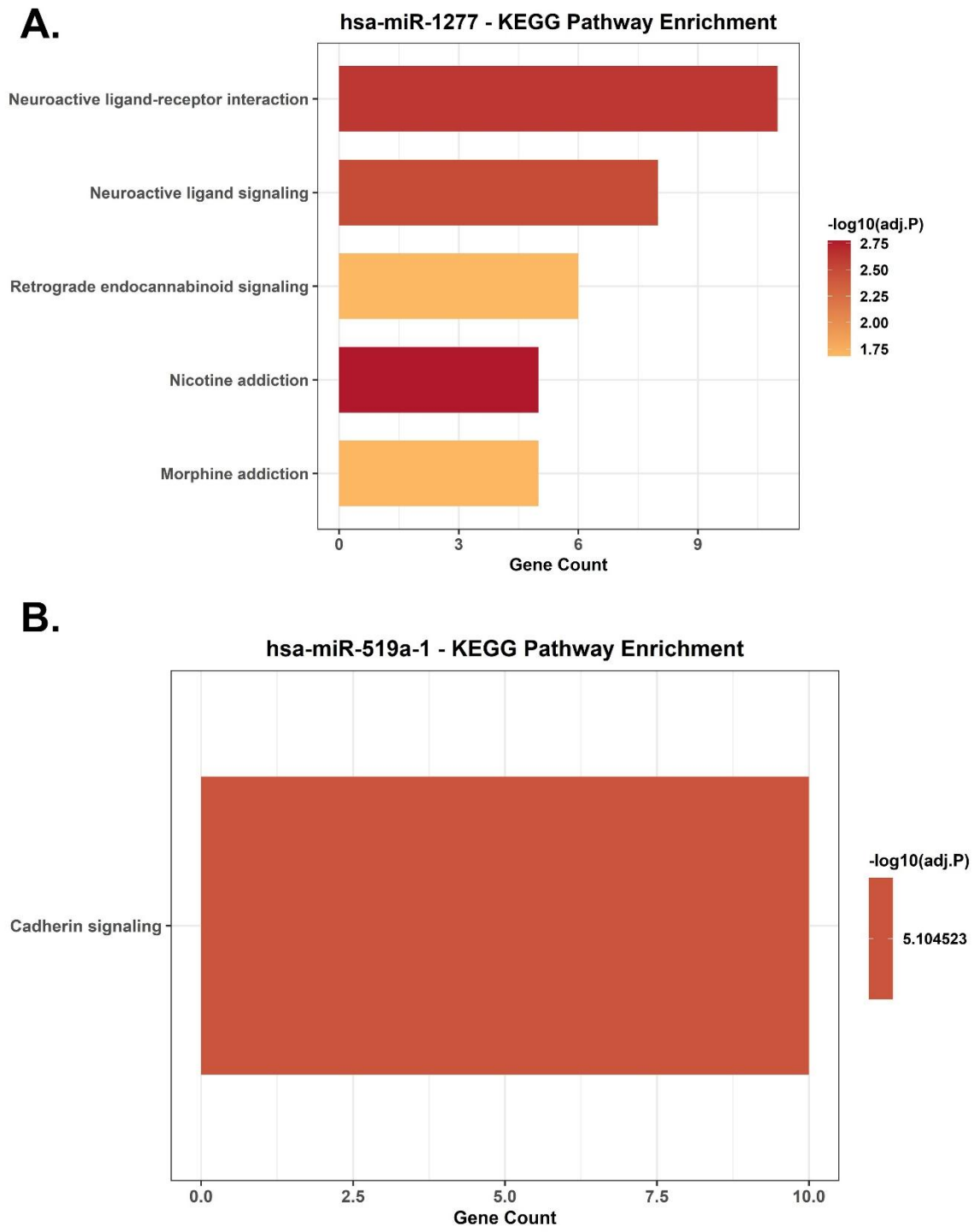


Figure 9. KEGG pathway enrichment analysis of candidate miRNA target genes. (A) KEGG pathway enrichment analysis of hsa-miR-1277-associated target genes demonstrating significant enrichment of neuroactive ligand–receptor interaction, retrograde endocannabinoid signaling, nicotine addiction, and morphine addiction pathways. **(B)** KEGG pathway enrichment analysis of hsa-miR-519a-1-associated target genes showing significant enrichment of the cadherin signaling pathway. Bar length represents the number of genes associated with each enriched pathway, while color intensity indicates enrichment significance represented as −log10 (adjusted p-value).

Consistent with these findings, KEGG pathway enrichment analysis of hsa-miR-1277-associated targets identified significant enrichment of neuroactive ligand–receptor interaction, retrograde endocannabinoid signaling, nicotine addiction, and morphine addiction pathways **(Figure 9A)**. Collectively, these pathways indicate strong association of the identified targets with neuronal signaling and receptor-mediated communication networks. Although several enriched pathways are classically linked to neuronal biology, accumulating evidence suggests that neurotransmission-associated signaling pathways also contribute to tumor cell communication, invasive behavior, and oncogenic signaling regulation in epithelial malignancies, including OSCC.

Additional GO Cellular Component and Molecular Function enrichment analyses further demonstrated significant enrichment of synaptic membrane-associated components, ion channel complexes, neurotransmitter receptor activity, gated ion channel activity, and potassium channel-related molecular functions among the candidate target genes. These findings further support the involvement of neuronal signaling and membrane-associated regulatory mechanisms in OSCC-associated miRNA dysregulation.

Similarly, GO enrichment analysis of hsa-miR-519a-1-associated target genes demonstrated significant enrichment of homophilic cell–cell adhesion and action potential-associated biological processes **(Figure 8B)**. KEGG pathway analysis further identified significant enrichment of the cadherin signaling pathway among hsa-miR-519a-1-associated targets **(Figure 9B)**, suggesting potential involvement in adhesion dysregulation and tumor-associated signaling processes. Notably, the enrichment profile was predominantly driven by protocadherin family-associated targets, supporting the interaction patterns observed in STRING network analysis.

In contrast, hsa-miR-301b-associated target genes did not demonstrate statistically significant GO Biological Process or KEGG pathway enrichment under the selected significance thresholds, potentially reflecting functional heterogeneity and limited pathway convergence among the analyzed targets.

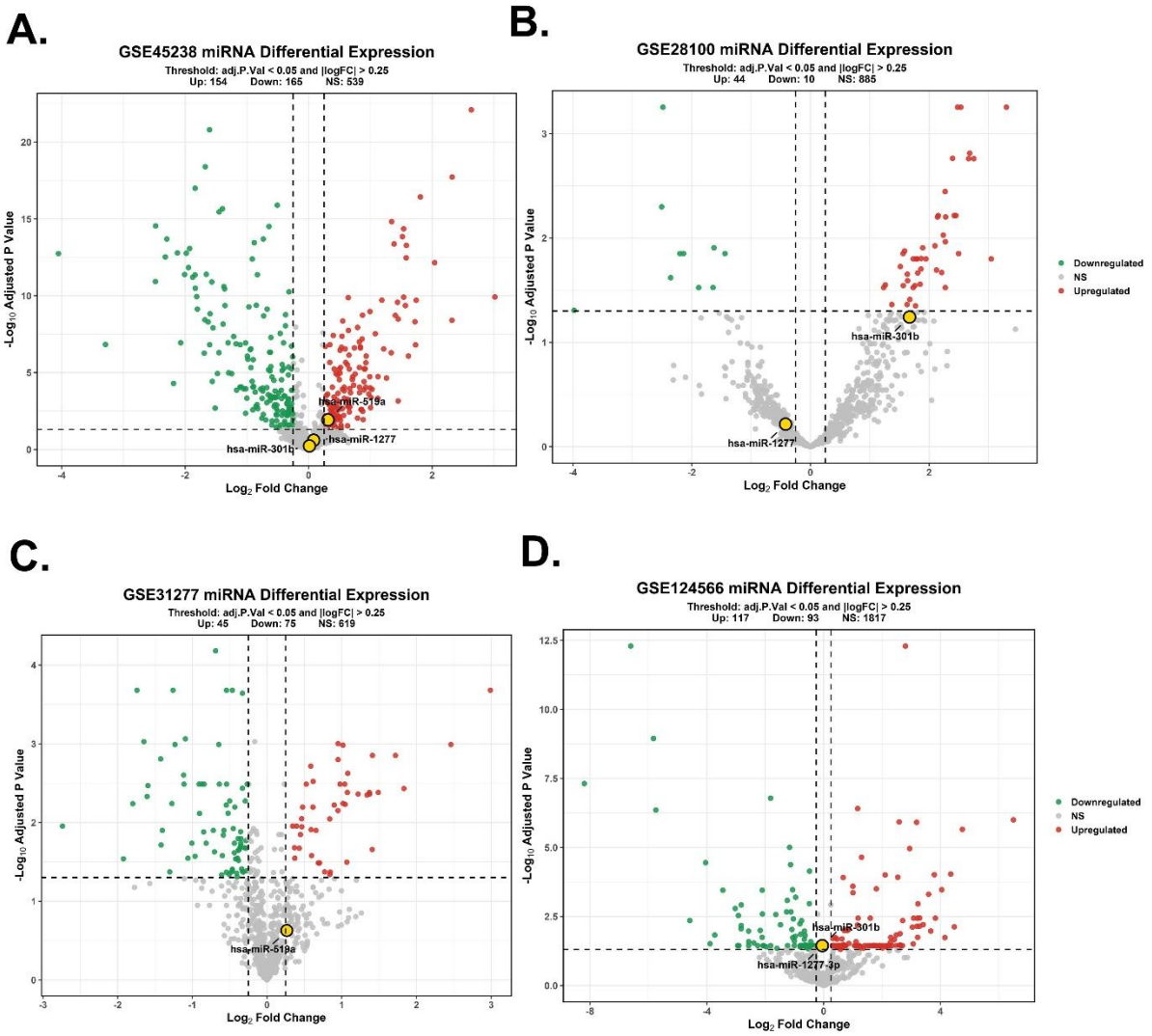


Figure 10. Differential miRNA expression patterns across independent GEO validation cohorts. Volcano plots illustrating differential miRNA expression profiles in (A) GSE45238, (B) GSE28100, (C) GSE31277, and (D) GSE124566. Differential expression analysis was performed using the limma framework with empirical Bayes moderation through GEO2R. Raw p-values were adjusted using the Benjamini–Hochberg false discovery rate (FDR) correction. Significantly dysregulated miRNAs were defined using adjusted p-value < 0.05 and $|\log_2 FC| \geq 0.25$. Upregulated miRNAs are shown in red, downregulated miRNAs in green, and non-significant miRNAs in gray. TCGA-derived candidate miRNAs are highlighted within each dataset to evaluate external validation consistency across independent cohorts.

7. **External Validation of Candidate miRNAs Across Independent GEO Cohorts:** To further validate the transcriptomic alterations identified in the TCGA-HNSC discovery cohort, differential expression analyses were performed across four independent GEO miRNA

expression datasets, including GSE45238, GSE28100, GSE31277, and GSE124566, using the limma framework with empirical Bayes moderation implemented through GEO2R **(Figure 10)**. Differentially expressed miRNAs were identified using Benjamini–Hochberg FDR-adjusted p-values < 0.05 and $|\log_2 FC| \geq 0.25$ **(Supplementary Document 10)**.

Volcano plot visualization demonstrated distinct differential expression distributions across the independent validation cohorts. Among the analyzed datasets, GSE45238 demonstrated the highest number of significantly dysregulated miRNAs, comprising 154 upregulated and 165 downregulated miRNAs. In contrast, GSE28100 exhibited comparatively fewer significant miRNAs, with 44 upregulated and 10 downregulated candidates. Similarly, GSE31277 identified 45 upregulated and 75 downregulated miRNAs, whereas GSE124566 demonstrated 117 upregulated and 93 downregulated miRNAs. The number of non-significant miRNAs varied substantially among datasets, reflecting differences in cohort composition, microarray platforms, normalization approaches, and expression variability.

Cross-validation of TCGA-derived candidate miRNAs across the GEO cohorts revealed heterogeneous expression patterns with only partial reproducibility among independent datasets. Several candidate miRNAs demonstrated directionally concordant dysregulation across multiple cohorts; however, statistical significance was not consistently retained in all datasets. This variability may be attributed to differences in sample size, platform-specific probe sensitivity, normalization methodologies, and the comparatively limited dynamic range associated with microarray-based miRNA profiling relative to sequencing-based datasets.

Notably, although certain candidate miRNAs demonstrated reduced statistical significance within individual GEO cohorts, their overall directional expression trends remained partially concordant with the TCGA-derived transcriptomic findings. Subsequent qRT-PCR validation similarly demonstrated heterogeneous expression patterns among clinical OSCC specimens, further supporting the substantial molecular heterogeneity characteristic of HNSCC. Collectively, the integrated analyses across TCGA discovery datasets, independent GEO validation cohorts, and experimental validation support the potential biological relevance of the identified miRNAs in OSCC progression.

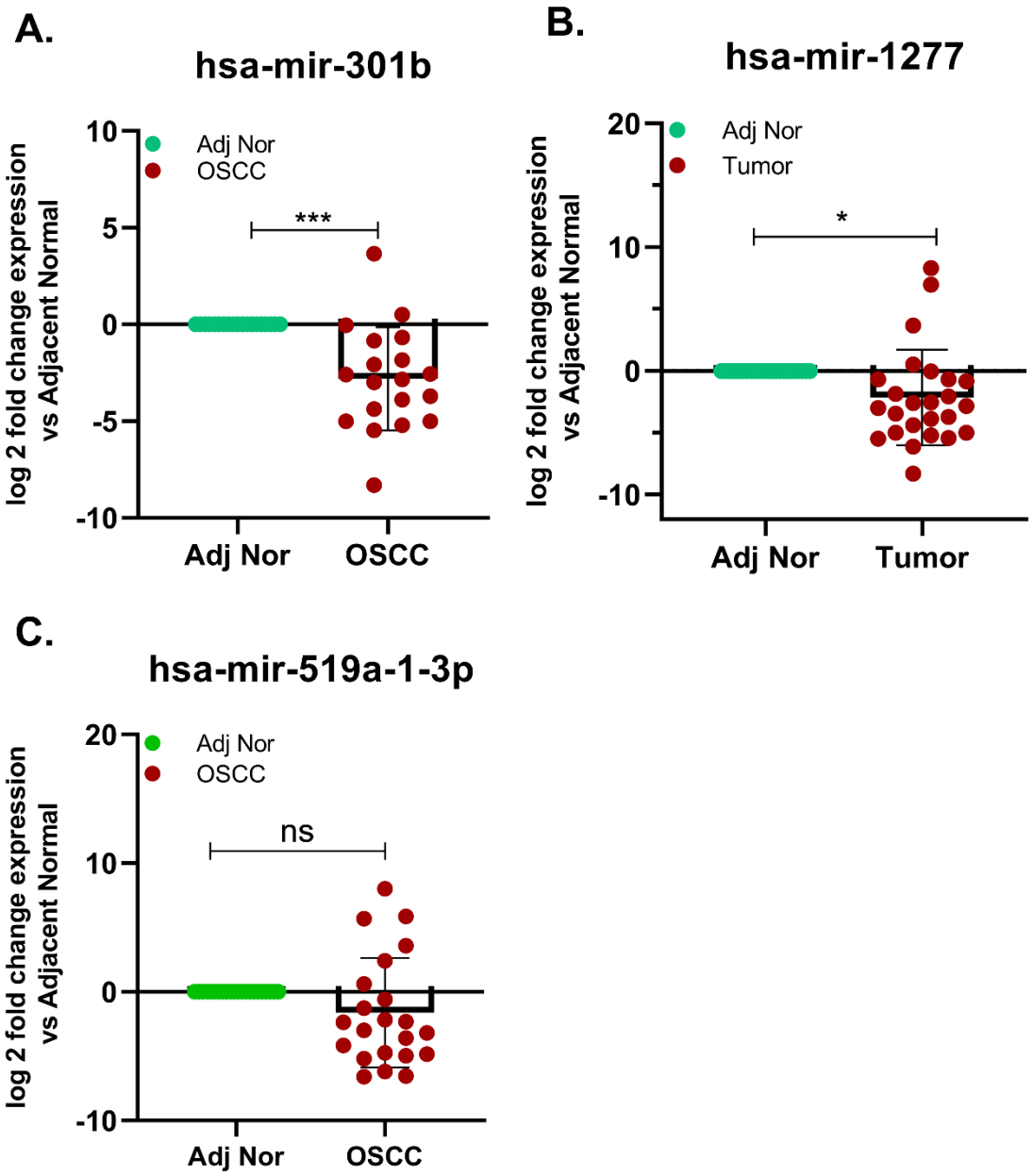


Figure 11. qRT-PCR-based validation of candidate miRNA dysregulation in OSCC clinical specimens. (A–C) Relative expression levels of hsa-miR-301b, hsa-miR-1277, and hsa-miR-519a-1 in oral squamous cell carcinoma (OSCC) tissues compared with adjacent normal oral mucosa (Adj Nor). Expression values are presented as $\log_2$ fold change relative to adjacent normal tissues. hsa-miR-301b demonstrated significant downregulation in OSCC tissues ($p = 0.0002$), while hsa-miR-1277 also showed significantly reduced expression ($p = 0.0105$). In contrast, hsa-miR-519a-1 did not exhibit statistically significant differential expression ($p = 0.0904$). Statistical analysis was performed using a one-sample t-test against a theoretical mean of zero. Horizontal bars represent mean ± SEM, and individual dots correspond to independent biological samples.

8. **qRT-PCR-Based Validation of Candidate miRNA Dysregulation in Indian OSCC Clinical Specimens:** To further validate the dysregulated expression patterns identified through TCGA-

HNSC transcriptomic analyses and independent GEO validation cohorts, quantitative real-time PCR (qRT-PCR) analysis was performed using OSCC clinical specimens and corresponding adjacent normal oral mucosa controls **(Figure 11)**. This experimental validation was undertaken to assess the reproducibility of the identified candidate miRNA alterations within an independent clinical cohort.

Relative miRNA expression levels were analyzed as $\log_2$ fold change values normalized against adjacent normal tissues, and statistical significance was evaluated using a one-sample t-test against a theoretical mean of zero **(Supplementary Document 11)**. Among the evaluated candidates, hsa-miR-301b demonstrated significant downregulation in OSCC tissues, with a mean $\log_2$ fold change of -2.805 ± 0.611 (95% CI: -4.089 to -1.520; $t = 4.587$, $p = 0.0002$) **(Figure 11A)**. Similarly, hsa-miR-1277 exhibited significantly reduced expression in OSCC specimens, with a mean $\log_2$ fold change of -2.148 ± 0.773 (95% CI: -3.744 to -0.552; $t = 2.777$, $p = 0.0105$) **(Figure 11B)**.

In contrast, hsa-miR-519a-1 did not demonstrate statistically significant differential expression between OSCC and adjacent normal tissues, despite showing an overall trend toward reduced expression (mean $\log_2$ fold change = -1.613 ± 0.909; 95% CI: -3.502 to 0.277; $t = 1.775$, $p = 0.0904$) **(Figure 11C)**.

Notably, substantial inter-sample variability was observed across individual clinical specimens, indicating pronounced molecular heterogeneity within the analyzed OSCC cohort. Although the expression patterns observed in qRT-PCR analyses were not entirely concordant with the TCGA-derived transcriptomic profiles, the collective findings obtained from computational analyses, external GEO validation cohorts, and experimental validation support the dysregulated expression of these candidate miRNAs in OSCC and further highlight the biological complexity associated with HNSCC progression.

DISCUSSION: MicroRNA dysregulation is increasingly recognized as a critical contributor to oral squamous cell carcinoma (OSCC) progression, influencing proliferation, invasion, epithelial–mesenchymal transition, stemness, and therapeutic response. Several studies have demonstrated that aberrant miRNA expression contributes substantially to the molecular heterogeneity of head and neck squamous cell carcinoma (HNSCC) **[28, 29]**. In the present study, integrative transcriptomic, bioinformatic, and experimental analyses identified hsa-miR-

1277, hsa-miR-301b, and hsa-miR-519a-1 as candidate miRNAs associated with OSCC-related molecular alterations.

Differential expression analysis of anatomically filtered HPV-negative TCGA-OSCC samples demonstrated widespread miRNA dysregulation, while ROC analysis supported the comparatively strong diagnostic potential of hsa-miR-301b and hsa-miR-1277. Previous investigations have similarly highlighted the diagnostic utility of miRNA signatures in distinguishing malignant oral lesions from normal tissues **[30]**. In contrast, Kaplan–Meier survival analysis demonstrated that none of the three miRNAs showed statistically significant associations with overall survival in the TCGA-HNSC cohort. Although minor fluctuations in survival trajectories were observed between high- and low-expression groups, these differences were not sustained across the cohort. These findings suggest that transcriptomic dysregulation alone may not necessarily translate into prognostic utility within heterogeneous HNSCC populations. Importantly, the TCGA-HNSC cohort comprises tumors from multiple anatomical subsites with distinct etiological and molecular characteristics, which may dilute OSCC-specific prognostic associations.

Integrative target prediction analysis combining TCGA-derived downregulated genes with multiple independent prediction algorithms identified several high-confidence candidate target genes potentially subjected to miRNA-mediated regulation. Among the evaluated candidates, hsa-miR-301b demonstrated the strongest and most consistent inverse correlations with multiple targets, including NR3C2, DLG2, AR, GABRB2, and PRKAA2. NR3C2 has previously been implicated in tumor suppressive signaling and cellular homeostasis in several malignancies, while PRKAA2-associated pathways are linked to metabolic regulation and cancer cell adaptation **[31]**. These findings collectively support the possibility that hsa-miR-301b may contribute to oncogenic progression through coordinated repression of multiple downstream regulators.

In contrast, hsa-miR-1277 demonstrated a comparatively selective inverse association, particularly with AGTR1. Angiotensin II receptor signaling has previously been associated with tumor proliferation, angiogenesis, inflammatory signaling, and metastasis in multiple cancers **[32]**. The observed inverse association between hsa-miR-1277 and AGTR1 therefore suggests a potentially relevant regulatory interaction in OSCC. Interestingly, HLF demonstrated a weak positive correlation with hsa-miR-1277 expression, emphasizing the

complexity of transcriptomic regulatory networks and suggesting that computationally predicted targets may not universally reflect direct inhibitory interactions in vivo.

Unlike hsa-miR-1277 and hsa-miR-301b, hsa-miR-519a-1 demonstrated comparatively weak and statistically non-significant transcript-level correlations with the selected target genes despite exhibiting a broader overlap profile during integrative target prediction analysis. Such discrepancies are biologically plausible and may arise from indirect regulation, translational repression without measurable transcript degradation, feedback regulatory loops, or context-dependent tumor heterogeneity. Importantly, miRNA-mediated regulation does not universally result in detectable transcript depletion within bulk transcriptomic datasets **[33]**.

STRING protein–protein interaction analysis further demonstrated moderate functional connectivity among several candidate targets. DLG2 emerged as a shared interaction node linking hsa-miR-1277 and hsa-miR-301b target networks, suggesting possible convergence on common signaling pathways. Functional enrichment analyses additionally revealed significant enrichment of neuronal signaling, synaptic regulation, ion channel activity, neurotransmitter receptor activity, and cadherin-associated pathways. Increasing evidence suggests that neuronal signaling pathways contribute substantially to tumor progression, neural crosstalk, invasion, and microenvironmental regulation in epithelial malignancies, including HNSCC **[34, 35]**. Similarly, enrichment of protocadherin and adhesion-associated pathways among hsa-miR-519a-1 targets suggests possible involvement in epithelial plasticity and cell adhesion-associated signaling dynamics.

In contrast, no statistically significant GO Biological Process or KEGG pathway enrichment was observed for hsa-miR-301b-associated targets despite strong inverse transcriptomic correlations. This apparent inconsistency likely reflects the relatively limited number of selected targets and their distribution across multiple unrelated biological processes, emphasizing that biologically relevant transcriptomic interactions may not necessarily converge into statistically overrepresented pathways.

External validation using independent GEO cohorts demonstrated partially concordant but heterogeneous expression patterns across datasets. Although several candidate miRNAs demonstrated directionally consistent dysregulation, statistical significance was not uniformly retained across all cohorts. Such variability may reflect differences in cohort composition, platform-specific probe sensitivity, normalization procedures, and the limited dynamic range associated with microarray-based profiling technologies relative to sequencing-based datasets.

Interestingly, qRT-PCR validation performed in Indian OSCC clinical specimens demonstrated expression trends that were not entirely concordant with TCGA-derived transcriptomic analyses. Rather than representing contradictory findings, these inconsistencies likely reflect population-specific molecular heterogeneity, ethnic variability, tobacco-associated molecular alterations, environmental exposure patterns, differences in tumor stage distribution, and relatively limited cohort size. Similar discrepancies between sequencing-based datasets and clinical validation cohorts have previously been reported in OSCC biomarker studies **[36]**.

Several limitations of the present study should be acknowledged. The analyses were primarily based on retrospective transcriptomic datasets and computational target prediction approaches, which cannot fully establish direct mechanistic regulation. Although qRT-PCR validation was performed, the clinical cohort size remained limited and lacked detailed molecular stratification. Furthermore, functional assays confirming direct miRNA–target interactions and downstream pathway modulation were not undertaken. Future investigations incorporating luciferase reporter assays, gain- and loss-of-function studies, transcriptomic perturbation analyses, and in vivo validation models will therefore be necessary to further establish the mechanistic contribution of these candidate miRNAs in OSCC progression.

Conclusion: The present study provides an integrative characterization of hsa-miR-1277, hsa-miR-301b, and hsa-miR-519a-1 in OSCC using TCGA transcriptomic profiling, GEO validation cohorts, target prediction analysis, Pearson correlation analysis, functional enrichment analysis, STRING interaction networks, and qRT-PCR-based clinical validation. Collectively, the findings indicate that these miRNAs may participate in distinct regulatory pathways involving neuronal signaling, receptor-mediated communication, metabolic regulation, synaptic signaling, and cell adhesion-associated processes in OSCC. Among the evaluated candidates, hsa-miR-301b and hsa-miR-1277 demonstrated comparatively stronger diagnostic relevance, whereas prognostic associations remained limited within the heterogeneous TCGA-HNSC cohort. Importantly, the inconsistencies observed across TCGA, GEO, and experimental validation cohorts further highlight the substantial molecular heterogeneity associated with OSCC and emphasize the complexity of miRNA-mediated regulatory mechanisms in head and neck cancer. Although additional mechanistic and large-scale clinical validation studies are required, the present findings expand the current understanding of miRNA-associated regulatory networks in OSCC and identify candidate

miRNAs that may warrant further investigation as potential biomarkers and therapeutic targets in head and neck cancer.

Abbreviations

HNSCC- Head and Neck Squamous Cell Carcinoma
OSCC – Oral Squamous Cell Carcinoma
miRNA – MicroRNA
TCGA – The Cancer Genome Atlas
GEO – Gene Expression Omnibus
ROC – Receiver Operating Characteristic
AUC – Area Under the Curve
GO – Gene Ontology
KEGG – Kyoto Encyclopedia of Genes and Genomes
PPI – Protein–Protein Interaction
STRING – Search Tool for the Retrieval of Interacting Genes/Proteins
qRT-PCR – Quantitative Real-Time Polymerase Chain Reaction
FDR – False Discovery Rate
DESeq2 – Differential Expression Sequencing 2
HPV – Human Papillomavirus
TCGA-HNSC – The Cancer Genome Atlas Head and Neck Squamous Cell Carcinoma Cohort
OS – Overall Survival
HR – Hazard Ratio
CI – Confidence Interval
BP – Biological Process
CC – Cellular Component
MF – Molecular Function
VST – Variance Stabilizing Transformation
GDC – Genomic Data Commons
RPM – Reads Per Million
SEM – Standard Error of Mean
PCA –Pearson Correlation Analysis

Statements & Declarations

Funding: Not applicable.

Competing Interest: The authors have no relevant financial or non-financial interests to disclose.

Authors Contribution: Both authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Antra. The first draft of the manuscript was written by Antra and Tandon V. The Study has been supervised by Tandon V. Both authors read and approved the final manuscript.

Ethics Approval and Consent to Participate

Human oral squamous cell carcinoma (OSCC) tumor specimens were obtained from 20 treatment-naïve patients undergoing surgical resection at Guru Teg Bahadur Hospital, following approval from the Institutional Ethics Committee (IECHR-2022-54-2-R1). Written informed consent was obtained from all participants in accordance with the Declaration of Helsinki.

Human and Animal Rights: Consent taken for Human specimens and Animal rights Not applicable.

Consent for publication: Not applicable.

Availability of Data and Materials: The raw data of the experiments are attached in the Supplementary documents. All other data supporting the findings of this study are available from the corresponding author upon reasonable request.

Clinical trial number: Not applicable.

Acknowledgements: Declared none.

References:

1. Jamali Z, Asl Aminabadi N, Attaran R, Pournagiazar F, Ghertasi Oskouei S, Ahmadpour F. MicroRNAs as prognostic molecular signatures in human head and neck squamous cell carcinoma: a systematic review and meta-analysis. Oral Oncol. 2015 Apr;51(4):321-31. doi: 10.1016/j.oraloncology.2015.01.008. Epub 2015 Feb 9. PMID: 25677760.

2. Chen D, Cabay RJ, Jin Y, Wang A, Lu Y, Shah-Khan M, Zhou X. MicroRNA Deregulations in Head and Neck Squamous Cell Carcinomas. J Oral Maxillofac Res. 2013 Apr 1;4(1):e2. doi: 10.5037/jomr.2013.4102. PMID: 24422025; PMCID: PMC3886106.

3. Babu JM, Prathibha R, Jijith VS, Hariharan R, Pillai MR. A miR-centric view of head and neck cancers. Biochim Biophys Acta. 2011 Aug;1816(1):67-72. doi: 10.1016/j.bbcan.2011.04.003. Epub 2011 Apr 28. PMID: 21549178.

4. Qiu K, Song Y, Rao Y, Liu Q, Cheng D, Pang W, Ren J, Zhao Y. Diagnostic and Prognostic Value of MicroRNAs in Metastasis and Recurrence of Head and Neck Squamous Cell Carcinoma: A Systematic Review and Meta-Analysis. Front Oncol. 2021 Sep 27;11:711171. doi: 10.3389/fonc.2021.711171. PMID: 34646767; PMCID: PMC8503605.

5. Vahabi M, Blandino G, Di Agostino S. MicroRNAs in head and neck squamous cell carcinoma: a possible challenge as biomarkers, determinants for the choice of therapy and targets for personalized molecular therapies. Transl Cancer Res. 2021 Jun;10(6):3090-3110. doi: 10.21037/tcr-20-2530. PMID: 35116619; PMCID: PMC8797920.

6. Yang CX, Sedhom W, Song J, Lu SL. The Role of MicroRNAs in Recurrence and Metastasis of Head and Neck Squamous Cell Carcinoma. Cancers (Basel). 2019 Mar 21;11(3):395. doi: 10.3390/cancers11030395. PMID: 30901831; PMCID: PMC6468798.

7. Hashmi AS, Gupta ND, Khan S, Ali SA, Kulsum O, Vellone V. Diagnostic and prognostic potential of salivary microRNA in oral and head and neck squamous cell

carcinomas: a systematic review. Int J Oral Maxillofac Surg. 2026 Jun;55(6):621-629. doi: 10.1016/j.ijom.2026.01.004. Epub 2026 Jan 23. PMID: 41580383.

8. Bhattacharjee B, Syeda AF, Rynjah D, Hussain SM, Chandra Bora S, Pegu P, Sahu RK, Khan J. Pharmacological impact of microRNAs in head and neck squamous cell carcinoma: Prevailing insights on molecular pathways, diagnosis, and nanomedicine treatment. Front Pharmacol. 2023 May 3;14:1174330. doi: 10.3389/fphar.2023.1174330. PMID: 37205904; PMCID: PMC10188950.

9. Cancer Genome Atlas Research Network; Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart JM. The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet. 2013 Oct;45(10):1113-20. doi: 10.1038/ng.2764. PMID: 24071849; PMCID: PMC3919969.

10. Goldman MJ, Craft B, Hastie M, Repečka K, McDade F, Kamath A, Banerjee A, Luo Y, Rogers D, Brooks AN, Zhu J, Haussler D. Visualizing and interpreting cancer genomics data via the Xena platform. Nat Biotechnol. 2020 Jun;38(6):675-678. doi: 10.1038/s41587-020-0546-8. PMID: 32444850; PMCID: PMC7386072.

11. Grossman RL, Heath AP, Ferretti V, Varmus HE, Lowy DR, Kibbe WA, Staudt LM. Toward a Shared Vision for Cancer Genomic Data. N Engl J Med. 2016 Sep 22;375(12):1109-12. doi: 10.1056/NEJMp1607591. PMID: 27653561; PMCID: PMC6309165.

12. Flynn R. Survival analysis. J Clin Nurs. 2012 Oct;21(19-20):2789-97. doi: 10.1111/j.1365-2702.2011.04023.x. Epub 2012 Aug 4. PMID: 22860755.

13. Therneau TM, Grambsch PM. *Modeling Survival Data: Extending the Cox Model*. New York (NY): Springer; 2000. 350 p.

14. Kassambara A, Kosinski M, Biecek P. survminer: Drawing Survival Curves using ‘ggplot2’. R package version 0.5.2.999. 2026.

15. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Müller M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011 Mar 17;12:77. doi: 10.1186/1471-2105-12-77. PMID: 21414208; PMCID: PMC3068975.

16. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 2002 Jan 1;30(1):207-10. doi: 10.1093/nar/30.1.207. PMID: 11752295; PMCID: PMC99122.

17. Wickham H. *ggplot2: Elegant Graphics for Data Analysis.* 2nd ed. Cham (CH): Springer; 2016. 260 p. doi:10.1007/978-3-319-24277-4.

18. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015 Apr 20;43(7):e47. doi: 10.1093/nar/gkv007. Epub 2015 Jan 20. PMID: 25605792; PMCID: PMC4402510.

19. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol. 2004;3:Article3. doi: 10.2202/1544-6115.1027. Epub 2004 Feb 12. PMID: 16646809.

20. Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. Elife. 2015 Aug 12;4:e05005. doi: 10.7554/eLife.05005. PMID: 26267216; PMCID: PMC4532895.

21. Chen Y, Wang X. miRDB: an online database for prediction of functional microRNA targets. Nucleic Acids Res. 2020 Jan 8;48(D1):D127-D131. doi: 10.1093/nar/gkz757. PMID: 31504780; PMCID: PMC6943051.

22. Sticht C, De La Torre C, Parveen A, Gretz N. miRWalk: An online resource for prediction of microRNA binding sites. PLoS One. 2018 Oct 18;13(10):e0206239. doi: 10.1371/journal.pone.0206239. PMID: 30335862; PMCID: PMC6193719.

23. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, Gable AL, Fang T, Doncheva NT, Pyysalo S, Bork P, Jensen LJ, von Mering C. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res. 2023 Jan 6;51(D1):D638-D646. doi: 10.1093/nar/gkac1000. PMID: 36370105; PMCID: PMC9825434.

24. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L, Fu X, Liu S, Bo X, Yu G. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation (Camb). 2021 Jul 1;2(3):100141. doi: 10.1016/j.xinn.2021.100141. PMID: 34557778; PMCID: PMC8454663.

25. Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M. KEGG: integrating viruses and cellular organisms. Nucleic Acids Res. 2021 Jan 8;49(D1):D545-D551. doi: 10.1093/nar/gkaa970. PMID: 33125081; PMCID: PMC7779016.

26. Sharbati-Tehrani S, Kutz-Lohroff B, Bergbauer R, Scholven J, Einspanier R. miR-Q: a novel quantitative RT-PCR approach for the expression profiling of small RNA molecules such as miRNAs in a complex sample. BMC Mol Biol. 2008 Apr 10;9:34. doi: 10.1186/1471-2199-9-34. PMID: 18400113; PMCID: PMC2374797.

27. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 2001 Dec;25(4):402-8. doi: 10.1006/meth.2001.1262. PMID: 11846609.

28. Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, Sweet-Cordero A, Ebert BL, Mak RH, Ferrando AA, Downing JR, Jacks T, Horvitz HR, Golub TR. MicroRNA expression profiles classify human cancers. Nature. 2005 Jun 9;435(7043):834-8. doi: 10.1038/nature03702. PMID: 15944708.

29. Childs G, Fazzari M, Kung G, Kawachi N, Brandwein-Gensler M, McLemore M, Chen Q, Burk RD, Smith RV, Prystowsky MB, Belbin TJ, Schlecht NF. Low-level expression of microRNAs let-7d and miR-205 are prognostic markers of head and neck squamous cell carcinoma. Am J Pathol. 2009 Mar;174(3):736-45. doi: 10.2353/ajpath.2009.080731. Epub 2009 Jan 29. PMID: 19179615; PMCID: PMC2665736.

30. Moreno LM, Mansilla MA, Bullard SA, Cooper ME, Busch TD, Machida J, Johnson MK, Brauer D, Krahn K, Daack-Hirsch S, L'heureux J, Valencia-Ramirez C, Rivera D, López AM, Moreno MA, Hing A, Lammer EJ, Jones M, Christensen K, Lie RT, Jugessur A, Wilcox AJ, Chines P, Pugh E, Doheny K, Arcos-Burgos M, Marazita ML, Murray JC, Lidral AC. FOXE1 association with both isolated cleft lip with or without cleft palate, and isolated cleft palate. Hum Mol Genet. 2009 Dec 15;18(24):4879-96. doi: 10.1093/hmg/ddp444. Epub 2009 Sep 24. PMID: 19779022; PMCID: PMC2778374.

31. Hardie DG. AMP-activated protein kinase: maintaining energy homeostasis at the cellular and whole-body levels. Annu Rev Nutr. 2014;34:31-55. doi: 10.1146/annurev-nutr-071812-161148. Epub 2014 May 15. PMID: 24850385; PMCID: PMC5693323.

32. George AJ, Thomas WG, Hannan RD. The renin-angiotensin system and cancer: old dog, new tricks. Nat Rev Cancer. 2010 Nov;10(11):745-59. doi: 10.1038/nrc2945. Epub 2010 Oct 22. PMID: 20966920.

33. Bartel DP. MicroRNAs: target recognition and regulatory functions. Cell. 2009 Jan 23;136(2):215-33. doi: 10.1016/j.cell.2009.01.002. PMID: 19167326; PMCID: PMC3794896.

34. Amit M, Takahashi H, Dragomir MP, Lindemann A, Gleber-Netto FO, Pickering CR, Anfossi S, Osman AA, Cai Y, Wang R, Knutsen E, Shimizu M, Ivan C, Rao X, Wang J, Silverman DA, Tam S, Zhao M, Caulin C, Zinger A, Tasciotti E, Dougherty PM, El-Naggar A, Calin GA, Myers JN. Loss of p53 drives neuron reprogramming in head and neck cancer. Nature. 2020 Feb;578(7795):449-454. doi: 10.1038/s41586-020-1996-3. Epub 2020 Feb 12. PMID: 32051587; PMCID: PMC9723538.

35. Zahalka AH, Frenette PS. Nerves in cancer. Nat Rev Cancer. 2020 Mar;20(3):143-157. doi: 10.1038/s41568-019-0237-2. Epub 2020 Jan 23. PMID: 31974491; PMCID: PMC7709871.

36. Courthod G, Franco P, Palermo L, Pisconti S, Numico G. The role of microRNA in head and neck cancer: current knowledge and perspectives. Molecules. 2014 May 5;19(5):5704-16. doi: 10.3390/molecules19055704. PMID: 24802984; PMCID: PMC6270891.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryDocument1TCGAHNSC.mirna.xls
- SupplementaryDocument2SurvivalAnalysis.xls
- SupplementaryDocument3ROC.xlsx
- SupplementaryDocument4hsamir301b5ptargets.xlsx
- SupplementaryDocument5hsamir519atargets.xlsx
- SupplementaryDocument6hsamir1277targets.xlsx
- SupplementaryDocument7PearsonCorrealtionAnalysis.xlsx
- SupplementaryDocument8StringNetwork.xlsx
- SupplementaryDocument9GOandKEGGEnrichment.xlsx
- SupplementaryDocument10GEODatasets.xlsx
- SupplementaryDocument11qPCRdatavalidation.xlsx