Biochemistry and Biophysics Reports 42 (2025) 102001

Contents lists available at ScienceDirect

Biochemistry and Biophysics Reports

journal homepage: www.elsevier.com/locate/bbrep

A comprehensive transcriptome based meta-analysis to unveil the aggression nexus of oral squamous cell carcinoma

Soujanya J. Vastrad [a], Ganesan Rajalekshmi Saraswathy [a,*], Jagadish B. Dasari [b], Gouri Nair [c], Ashok Madarakhandi [d], Dominic Augustine [e], S.V. Sowmya [e]

[a] Department of Pharmacy Practice, Faculty of Pharmacy, M.S. Ramaiah University of Applied Sciences, New BEL Road, M.S.R. Nagar, Bengaluru, India
[b] Department of Biology, University of Rome Tor Vergata, Rome, Italy
[c] Department of Pharmacology, Faculty of Pharmacy, M. S. Ramaiah University of Applied Sciences, Bangalore, Karnataka, India
[d] Department of Pharmaceutical Chemistry, KLE College of Pharmacy, (A Constituent Unit of KAHER-Belagavi), 2nd Block, Rajajinagar, Bangalore, India
[e] Department of Oral Pathology and Microbiology, Faculty of Dental Sciences, M.S. Ramaiah University of Applied Sciences, New BEL Road, M.S.R. Nagar, Bengaluru, India

ARTICLE INFO

Keywords:
Oral cancer
Nodal metastasis
Recurrence
Transcriptomic meta-analysis
Biomarkers
Genomics
Squamous cell carcinoma

ABSTRACT

Lymph node metastasis in oral cancer (OC) complicates management due to its aggressive nature and high risk of recurrence, underscoring the need for biomarkers for early detection and targeted therapies. However, the drivers of this aggressive phenotype remain unclear due to the variability in gene expression patterns. To address this, an integrative meta-analysis of six publicly available transcriptomic profiles, categorized by lymph nodal status, is conducted. Key determinants of disease progression are identified through functional characterization and the TopConfects ranking approach of nodal associated differentially expressed genes (DEGs). To explore the critical nexus between lymph node metastasis and OC recurrence, significant metastatic genes were cross-analysed with literature-derived genes exhibiting aberrant methylation patterns in OC recurrence. Their clinical relevance and expression patterns were then validated in an external dataset from the TCGA head and neck cancer cohort. The analysis identified elevated expression of genes involved in extracellular matrix remodelling and immune response, while the expression of genes related to cellular differentiation and barrier functions was reduced, driving the transition to nodal positivity. The highest-ranked gene, MMP1, showed a log-fold change (LFC) of 4.946 (95 % CI: 3.71, 6.18) in nodal-negative samples, which increased to 5.899 (95 % CI: 4.80, 6.99) in nodal-positive samples, indicating consistent elevation across disease stages. In contrast, TMPRSS11B was significantly downregulated, with an LFC of −5.512 (95 % CI: −6.63, −4.38) in nodal-negative samples and −5.898 (95 % CI: −7.15, −4.64) in nodal-positive samples. Furthermore, MEIS1, down-regulated in nodal-positive status, was found to exhibit hypermethylation at CpG sites associated with OC recurrence. This study represents the first transcriptomic meta-analysis to explore the intersection of lymph node metastasis and OC recurrence, identifying MEIS1 as a potential key contributor. These comprehensive insights into disease trajectories offer potential biomarkers and therapeutic targets for future treatment strategies.

1. Introduction

Oral squamous cell carcinoma (OSCC) is the most common type of malignant neoplasm of head and neck squamous cell carcinoma (HNSC), with around 90 % of tumours developing within the epithelial lining of the oral cavity [1]. In its early, localized stages, OSCC typically has a positive prognosis, with a high 5-year survival rate. However, this outlook worsens significantly when the disease spreads to neighbouring lymph nodes, resulting in a considerable reduction in the survival rate. The presence of lymph node metastasis is a critical, independent prognostic factor that affects patient outcomes. OSCC tends to metastasize to the cervical lymph node at an early stage, further complicating the therapeutic decisions [2,3]. More than 50 % of OSCC patients often present with locally advanced disease at the time of initial diagnosis, necessitating an interdisciplinary treatment approach involving surgery, radiotherapy, and chemotherapy [1]. Surgery remains the mainstay of treatment, with studies indicating improved survival outcomes compared to primary radiotherapy [4,5]. However, treatment-related

* Corresponding author.
E-mail address: saraswathypradish@gmail.com (G.R. Saraswathy).

https://doi.org/10.1016/j.bbrep.2025.102001
Received 5 January 2025; Received in revised form 21 March 2025; Accepted 31 March 2025

disability is a concern, and many patients experience disease recurrence or distant metastasis. In scenarios of recurrent or metastatic OSCC, the standard treatment involves palliative systemic therapy. This approach typically results in a median survival duration of 6–9 months and a 1-year survival rate ranging from 20 % to 40 % with chemotherapy as the sole intervention [6].

In the realm of OSCC treatment strategies, the clinical TNM classification system governs the treatment planning, anticipating smaller, less invasive tumours without metastasis to demonstrate a favourable prognosis. However, tumours with identical staging can exhibit different growth patterns and clinical behaviours, complicating prognostic predictions. Despite numerous histopathological factors being identified as prognostic markers, none are definitive. Also, accurately assessing tumour depth often poses specific challenges [7,8].

Therefore, there is an acute need to improve outcomes in this patient group. Considering the magnitude of the intricate pathological cascades underlying disease progression and invasiveness, alongside the dearth of robust strategies in addressing the same, mandates a comprehensive understanding of the disease panorama. Thus, unearthing reliable OSCC gene signatures or molecular biomarkers would serve as a valuable diagnostic or prognostic tool in the management of the disease.

The clinical and therapeutic significance of molecular subtyping, including DNA, RNA, and protein profiling, has been especially prominent in cancer research [9]. Comprehensive transcriptome (RNA) analysis has emerged as a highly effective method for examining broad gene expression alterations, thereby facilitating insights into disease origin, metastatic development, and guiding diagnostic and therapeutic strategies [10]. Despite its potential, the application of transcriptomics in clinical trials is often limited by small sample sizes, leading to challenges in analysing high-dimensional data where the number of genes far exceeds the available samples, making the analysis difficult and less reliable [11–13].

Transcriptomic studies in OSCC research, involving preclinical models and clinical specimens, are commonly utilized to elucidate disease mechanisms, identify novel biomarkers and potential therapeutic targets. The most appropriate samples available for these analyses are biopsies of oral tumour tissues, often obtained during diagnostic procedures such as surgical resections. These tissue samples serve as primary sources for histopathological assessment of tumour characteristics and are also pivotal for molecular studies [14]. Various factors such as biopsy site, distance from the tumour margin, presence of malignant alterations, and tissue sample integrity can significantly influence transcriptomic analysis outcomes [15].

Previously, meta-analyses of data from multiple OSCC studies have been undertaken using various methodologies to explore changes in cellular physiology, expression of heat-shock proteins and E-cadherin, and exosomal protein-based biomarkers [16–19].

The current study focuses on 1) The Discernment of genes involved in the progression of OSCC, and 2) Unveiling the factors related to disease aggressiveness that contribute to both recurrence and lymph node metastasis. Herein, a random effects meta-analysis of publicly available microarray datasets derived from oral cancer biopsies was performed to retrieve set of significant genes involved in OSCC lymph node metastasis. Subsequently, these genes were mapped to the disease recurrence to disinter the aggression nexus (Fig. 1).

2. Methods

All the bioinformatics and statistical analyses were performed utilizing R software v.4.3.1 [21].

2.1. Screening of omics data to retrieve eligible OSCC transcriptomic profiles

OSCC transcriptomic profiles were obtained from the NCBI, Gene Expression Omnibus (GEO) [22], a public repository database. A systematic search of all the reposited transcriptomic profiles (datasets): 2002–2023 was undertaken between January and April of 2023, in accordance with the conventional procedure for meta-analyses [23]. The structured Boolean query for this search included a combination of various MeSH terms: (((((((Mouth Neoplasm) OR Oral Neoplasm) OR Oral Cancer) OR Oral Squamous Cell Carcinoma) OR Mouth Cancer) OR

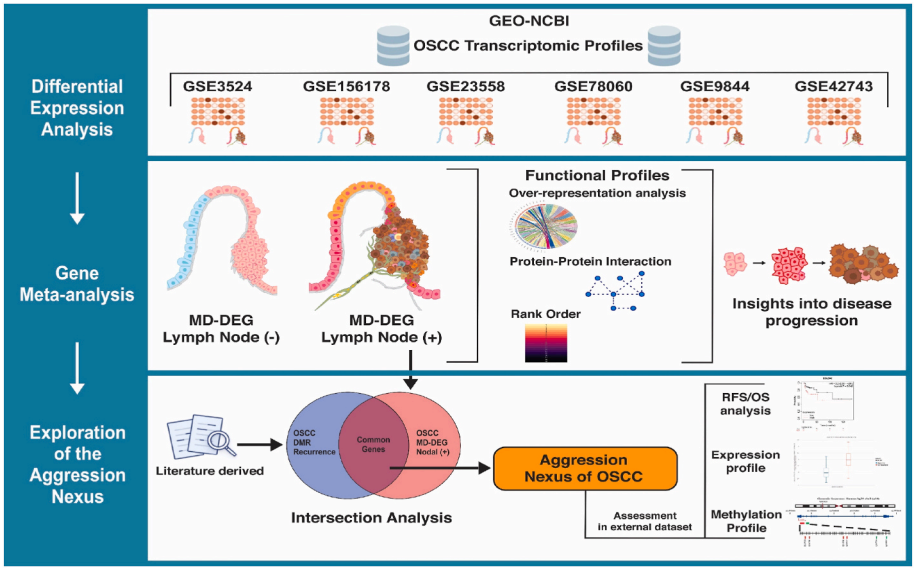

Fig. 1. Methodological framework: The selected transcriptomic profiles of Oral squamous cell carcinoma (OSCC) were systematically retrieved from the GEO-NCBI data repositories. Following data exploration and preprocessing, differential expression analysis was performed on the selected OSCC datasets, categorized by lymph node statuses (negative and positive). Subsequently, a transcriptome based meta-analysis was conducted, integrating the results of individual nodal negative (−) and positive (+) statuses. The significant differentially expressed genes (DEGs) identified via the meta-analysis (MD-DEG lymph node (−) and MD-DEG lymph node (+)) were further subjected to functional, protein-protein interaction analysis and ranking to gain insights into disease progression. To investigate the genes driving the aggressive behaviour of OSCC, involving lymph node metastasis and recurrence, the intersection of literature-derived genes [20] exhibiting differentially methylated regions (DMRs) linked to recurrence and significant OSCC MD-DEG lymph node (+) was analysed. The resulting candidates within this aggression nexus were evaluated using the independent head and neck squamous cell carcinoma (HNSC) dataset.

Tongue Neoplasms) OR Tongue Cancer) OR Gingival Neoplasms) OR OSCC.

The following inclusion criteria were applied.

- Studies involving both non-OSCC (control) and OSCC (case) samples
- Sample size >3 for both the case and control groups
- Transcriptome Analysis of tissue samples from Homo sapiens
- RNA extracted directly from oral anatomical sub-sites as defined by the American Joint Committee on Cancer (AJCC)
- OSCC samples with nodal status information

Datasets with the following criteria were excluded.

- Datasets containing drug-treated samples
- Studies examining gene regulation mechanisms
- Mutational studies

Subsequently, gene expression data from the datasets containing filtered samples from specific OSCC anatomical sites along with nodal status information were retrieved using the R package GEOquery [24].

2.2. Differential gene expression analysis of individual OSCC transcriptomic profiles with Varying nodal status

The datasets were individually pre-processed followed by transcriptomic analyses of each dataset. Differential gene expression analysis was performed between i) normal non-OSCC control group vs OSCC case groups without lymph node metastasis and ii) normal non-OSCC control group vs OSCC case groups with lymph node metastasis.

The data preprocessing included standardizing the nomenclature of clinical variables across studies. Raw microarray data were retrieved from the GEO database, assessed, log2-transformed, and normalized. For RNA-seq datasets, low-expressed raw counts were filtered, and the 'voom' function from the limma package was applied to convert the raw counts into log-CPM values with precision weights for downstream modelling. Probe sets were mapped to Ensembl gene IDs and annotated to HUGO gene symbols using the biomaRt R package [25, 26]. Duplicate mappings were resolved by calculating the median expression. The datasets were explored using unsupervised clustering to identify expression patterns across samples and genes and to assess the presence of batch effects if any.

Differential expression analysis was conducted using the limma R package [27–29]. The P-values were adjusted for multiple testing using the Benjamini-Hochberg procedure, with statistical significance defined as adj.p-value <0.05 and a log2 fold change (LFC) > 1.5 (up or down-regulated).

2.3. Meta-analysis of OSCC specific DEGs with Varying nodal status

To perform a meta-analysis of screened transcriptomic profiles, we systematically integrated the results of OSCC nodal status-based differentially expressed genes (DEGs) derived from individual datasets into two separate meta-analysis groups: OSCC-Nodal (−) and OSCC-Nodal (+). Meta-analysis was conducted using R package metafor [30], with the application of the DerSimonian and Laird random-effects model to accommodate individual study heterogeneity [31]. This methodology addresses variability across individual studies by assigning greater weights to studies with lower variability during the computation of meta-analysis results. Consequently, this approach emphasizes the most robust genes identified across the studies. To evaluate the overall similarities in DEGs across individual datasets and the meta-analysis, PCA principal component analysis (PCA) was utilized to reduce the data dimensionality based on LFC values, using the R base package stats and pcaExplorer [32] for all the genes. Additionally, forest plots were generated to compare the mean LFC values to determine the impact of individual datasets on the meta-analysis outcomes.

The DEGs that were observed consistently across at least 80 % of the analysed datasets, with an LFC >1.5 (up- or down-regulated) and exhibiting an adj.p < 0.05, were considered significant in the meta-analysis.

To interpret the genes associated with disease progression, significant meta-analysis-derived differentially expressed genes (MD-DEGs) were ranked using the TopConfects approach. The top 10 highly ranked MD-DEGs (up- and down-regulated) for each nodal status (positive and negative) were thus utilized to characterize their relation to the disease progression [33].

2.4. Analysis of nodal status based functional signatures

The set of nodal status associated DEGs identified through meta-analysis was scrutinized to elucidate their significance in attributing to functional differences between nodal negative and positive. Over-representation analysis (ORA) was carried out to examine the enriched biological functions and pathways associated with upregulated and downregulated genes corresponding to individual nodal statuses using the R package clusterProfiler [34]. P-values and adj.p-val (<0.05) were determined for each GO term [35] and KEGG pathway [36].

2.5. Protein-protein interaction (PPI) analysis to identify hub genes crucial in OSCC progression

The PPI network of significant nodal status-associated MD-DEGs was constructed using the STRING web tool [37] and subsequently was visualized with Cytoscape software (version-3.10.1) [38]. MD-DEGs that were a part of the network displaying potential association with nodal-negative status; common to both nodal positive and negative statuses; and association with nodal-positive status were designated as Nodal (−) PPI; Nodal (±) PPI and Nodal (+) PPI subnetworks, respectively.

2.6. Identification and evaluation of potential genes underpinning the aggression nexus of OSCC

To elucidate the aggression nexus in oral cancer, we investigated the potential existence of epigenetic modifications in nodal metastatic genes that may influence disease recurrence. For this, an intersection analysis was performed between the significant MD-DEGs identified in the current analysis of the nodal-positive status and literature-derived genes exhibiting aberrant differentially methylated regions (DMR) associated with recurrence, from the published reports by Sorroche et al. [20]. The resulting common genes were analysed for relapse-free survival (RFS) and overall survival (OS) in independent HNSC cohort, using the Kaplan-Meier plotter database [39].

This was followed by a comprehensive evaluation of expression and methylation profiles in the OSCC dataset derived from GDC TCGA-HNSC cohort using UCSC Xena, a web browser database [40].

2.7. Correlations between methylation patterns and mRNA expression level of the shortlisted gene

To explore the relationship between DNA methylation pattern and gene expression, MEXPRESS was utilized, a web-based application designed to integrate and visualize expression and methylation data [41, 42]. Pearson correlation analysis was employed to compare the methylation status with corresponding mRNA expression levels [41]. The correlation between methylation pattern and mRNA expression level for the shortlisted gene was examined in the TCGA-HNSC dataset comprising 612 human samples.

Additionally, the genomic structure was obtained from the UCSC genome browser [43] including a chromosomal ideogram. The transcriptional start site (TSS), promoter, and enhancer regulatory sites were referenced from the NCBI database. The location of the TSS was also

confirmed using the KERO database [44]. This analysis allowed for precise mapping of the methylation sites in relation to the regulatory elements, providing insights into the potential impact of methylation patterns on gene expression.

3. Results

3.1. Screening of OSCC transcriptomic profiles

The systematic review of the NCBI GEO database revealed 393 OSCC transcriptomic profiles, which upon removal of duplication resulted in 387 series records. Following the application of inclusion and exclusion criteria, six qualified microarray datasets (GSE3524, GSE156178, GSE23558, GSE78060, GSE9844, and GSE42743) were selected for the analysis (Fig. 2). These datasets were sourced from various institutions and microarray platforms (Table 1). The range of OSCC samples in each dataset, without and with lymph node metastasis, varied from 5 to 29 (median 12.5) and from 6 to 42 (median 12), respectively. The pooled data comprised a total of 181 OSCC cases (81 without, and 100 with lymph nodal spread, respectively).

3.2. OSCC specific DEGs with Varying nodal status

The datasets were subjected to exploratory and pre-processing steps to ensure their compatibility for subsequent analysis (Supplementary Figs. 8a–f). The differential expression results across studies exhibited a wide range of significant genes (adj.p-value <0.05, LFC >1.5) showing up-regulation and down-regulation. In the case of OSCC-Nodal (−) status: the median count of up-regulated genes was 40, with a range extending from 36 (GSE9844) to 269 (GSE156178), while the median count of down-regulated genes was 65 ranging from 28 (GSE9844) to 760 (GSE156178). Similarly in the case of OSCC-Nodal (+) status: the median count of up-regulated genes was 110, with a range extending from 68 (GSE3524) to 817 (GSE156178), while the median count of down-regulated genes was 234, ranging from 122 (GSE3524) to 1068 (GSE156178) (Table 2). At this stage, the differential analysis of GSE78060 for nodal negative status did not reveal any significant gene expression changes.

3.3. OSCC specific meta-analysis of DEGs with Varying nodal status

A random effect meta-analysis was performed across five datasets in the OSCC-Nodal (−) group and six datasets in the OSCC-Nodal (+) group. The meta-analysis of OSCC-Nodal (−) group revealed significant 3480 DEGs in 4 or more datasets. Of these, 40 up- and 65 down-regulated genes met the criteria LFC >1.5 and adj.p-value <0.05. Similarly, the meta-analysis of DEG of OSCC-Nodal (+) group revealed 5153 genes up- or down-regulated genes in 5 or more datasets. Of these, significant 76 up- and 95 down-regulated genes met the criteria of LFC >1.5 and adj.p-value <0.05 (Supplementary Tables 2 and 3). The complete list of MD-DEGs for negative and positive nodal status can be found in Supplementary Tables 9 and 10.

The PCA analysis performed to evaluate the overall DEG similarities between the individual datasets and meta-analysis conducted revealed: 1) In the case of OSCC-Nodal (−), the first principal component (PC1) accounted for 63.77 % of the total variance, while an additional 21.59 % of the variance was observed in the second principal component (PC2) 2) In the case of OSCC-Nodal (+), the first principal component (PC1) accounted for 65.85 % of the total variance, while an additional 15.53 % of the variance was observed in the second principal component (PC2).

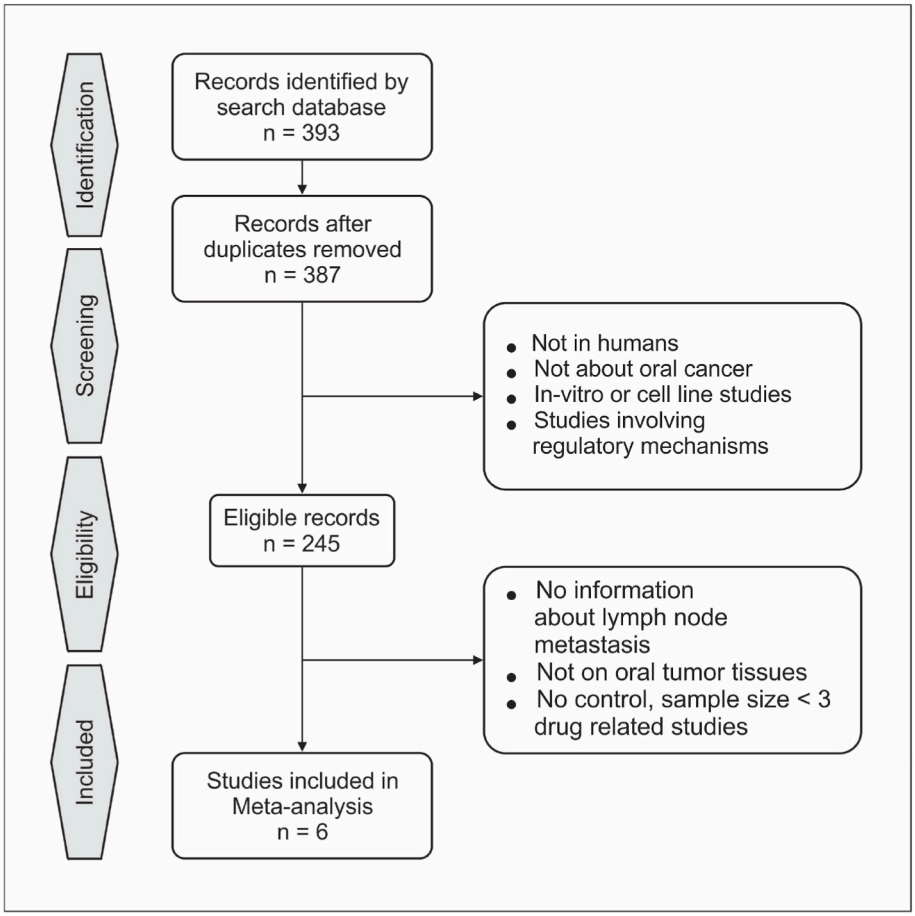

Fig. 2. Screening of omics data to retrieve eligible OSCC transcriptomic profiles for Meta-analysis.

Table 1
Gene Expression Omnibus (GEO) datasets used for the Meta-analysis.

GEO series Accession number	Publication PMID	Total OSCC case samples (N)	Non-OSCC control samples (N)	OSCC-Nodal (−) (N)	OSCC-Nodal (+) (N)	Biopsy location	Array platform
GSE3524	15381369	14	4	7	7	Floor of mouth, Tongue, Vestibule, Maxilla, Mandible	Affymetrix Human Genome U133A Array
GSE156178	32895418	17	5	11	6	Gingiva, Tongue	Illumina HiSeq 2500 (Homo sapiens)
GSE23558	22072328	27	4: Independent 1: Pooled sample from 9 other healthy donors	14	13	Gingivobuccal complex	Agilent-014850 Whole Human Genome Microarray 4x44K G4112F
GSE78060	28977904	26	4	5	21	Tongue	Affymetrix Human Genome U133 Plus 2.0 Array
GSE9844	18254958	26	12	15	11	Tongue	Affymetrix Human Genome U133 Plus 2.0 Array
GSE42743	23319825	71	29	29	42	Oral Cavity	Affymetrix Human Genome U133 Plus 2.0 Array

PMID: PubMed unique identifier; N: number of samples; OSCC-Nodal (−): OSCC case group without lymph node metastasis; OSCC-Nodal (+): OSCC case group with lymph node metastasis.

Table 2
Summary of differential gene expression analysis: OSCC Nodal status-based comparison.

GEO series Accession number	OSCC-Nodal (−) significant DEGs			OSCC-Nodal (+) significant DEGs		
	Up-regulated genes	Down-regulated genes	Total	Up-regulated genes	Down-regulated genes	Total
GSE3524	40	33	73	68	122	190
GSE156178	269	760	1029	817	1068	1885
GSE23558	169	375	544	136	314	450
GSE78060	–	–	–	320	287	607
GSE9844	36	28	64	84	143	227
GSE42743	40	65	105	83	181	264

OSCC-Nodal (−): OSCC case group without lymph node metastasis; OSCC-Nodal (+): OSCC case group with lymph node metastasis; Significant DEGs (Differentially expressed genes): genes satisfying log2 fold change >1.5 and adj.p-value <0.05.

At this point, the dataset GSE156178 exhibited a greater deviation along the PC1 axis in both the nodal negative and positive status. To assess the robustness of the analysis, a validation check was conducted. PCA was reanalysed excluding GSE156178, and the resulting Euclidean distance matrices before and after exclusion were visualized using an MDS plot. The overall structure was maintained (Supplementary Fig. 9a and b).

The meta-analysis dataset was approximately centred on the PC2 axis in the PCA of both OSCC-Nodal (−) and OSCC-Nodal (+), thus confirming the 'average' representation of the individual datasets in the meta-analysis (Fig. 3).

The significant DEGs derived from meta-analysis of negative and positive nodal statuses were assigned ranks based on TopConfects approach. Specifically, in the lymph node-negative status, the highly ranked (R) upregulated genes among the top 10 positions included CXCL10 (R-18), AIM2 (R-20), PARP12 (R-27), and CXCL11 (R-29). Similarly, MMP10 (R-5), MMP3 (R-15), INHBA (R-17), and GSDME (R-25) were observed in lymph node-positive status. Additionally, MMP1, PTHLH, ISG15, MMP12, IFI6, and BST2 were observed in both positive and negative nodal statuses. In the case of highly ranked downregulated genes, RHCG (R-12) was observed in lymph node-negative status, while FAM3B (R-7) was observed in lymph node-positive status. Furthermore, MAL, TMPRSS11B, CRNN, KRT13, KRT4, TGM3, ENDOU, SPINK5, and SPINK7 were detected in both positive and negative nodal statuses (Table 3).

To assess the influence of each dataset on the results of the meta-analysis, forest plots were generated to compare the average LFC and

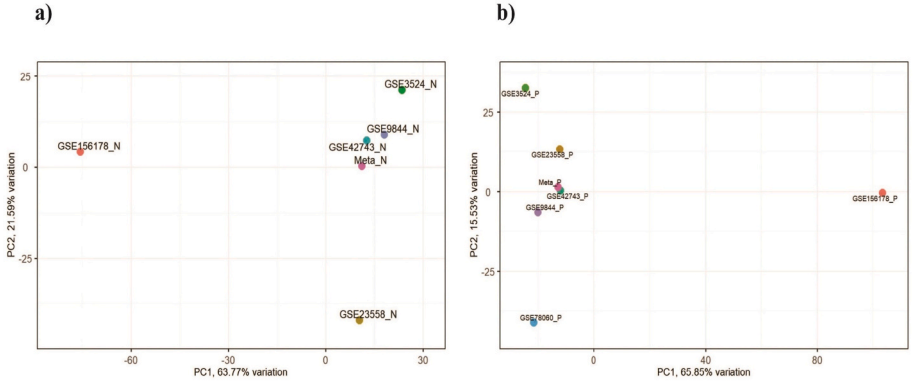

Fig. 3. Principal component analysis (PCA) of the datasets, based on log2 fold change of all the differentially expressed genes in Oral squamous cell carcinoma (OSCC). The plot shows the first principal component (PC1) and the second principal component (PC2) with the percentages indicating the variance explained by each component. a) OSCC-Nodal (−) Meta-analysis, Meta-N represents the meta-analysis of nodal negative status of datasets b) OSCC-Nodal (+) Meta-analysis, Meta-P represents the meta-analysis of nodal positive status of datasets.

Table 3
Top 10 up-regulated and down-regulated differentially expressed genes associated with nodal-negative and nodal-positive status, as identified by an OSCC meta-analysis using the TopConfects ranking approach.

Up-regulated expression			Down-regulated expression		
Gene Symbol	LFC (95 % CI)	adj.p-value	Gene Symbol	LFC (95 % CI)	adj.p-value
Rank order of OSCC-Nodal (−) genes derived from Meta-analysis					
MMP1	4.946 (3.71, 6.18)	1.91E-04	TMPRSS11B	−5.512 (−6.63, −4.38)	3.40E-09
ISG15	2.473 (2.03, 2.90)	7.67E-05	CRNN	−4.838 (5.91, −3.75)	6.01E-07
BST2	2.306 (1.59, 3.01)	4.84E-03	MAL	−5.437 (−7.02, −3.85)	1.20E-09
CXCL10	2.707 (1.63, 3.78)	1.72E-05	ENDOU	−3.64 (−4.38, −2.89)	9.10E-09
PTHLH	2.631 (1.60, 3.65)	5.05E-08	KRT13	−4.069 (−5.18, −2.95)	1.00E-09
AIM2	1.848 1.47, 2.22)	2.08E-05	TGM3	−4.04 (−5.24, −2.83)	2.00E-10
MMP12	2.262 (1.46, 3.05)	1.19E-03	SPINK7	−3.533 (−4.75, −2.30)	7.10E-09
PARP12	1.664 (1.38, 1.94)	2.11E-04	SPINK5	−3.003 (−4.02, −1.98)	1.27E-07
CXCL11	2.066 (1.41, 2.71)	8.03E-07	KRT4	−3.949 (−5.69, −2.2)	3.00E-03
IFI6	2.138 (1.41, 2.86)	5.43E-05	RHCG	−2.594 (−3.32, −1.86)	6.07E-05
Rank order of OSCC-Nodal (+) genes derived from Meta-analysis					
MMP1	5.899 (4.80, 6.99)	7.28E-05	MAL	−5.815 (−6.89, −4.73)	4.21E-05
MMP10	4.176 (3.52, 4.82)	3.10E-05	TMPRSS11B	−5.898 (−7.15, −4.64)	3.29E-06
PTHLH	3.56 (2.81, 4.30)	3.19E-06	CRNN	−5.487 (−6.53, −4.44)	4.39E-07
ISG15	2.827 (2.24, 3.40)	3.19E-05	KRT13	−4.929 (−6.35, −3.50)	2.19E-05
MMP12	3.091 (2.29, 3.88)	3.28E-06	FAM3B	−4.564 (−5.99, −3.13)	4.71E-06
MMP3	3.759 (2.42, 5.09)	9.39E-06	KRT4	−4.356 (−5.70, −3.00)	1.90E-05
INHBA	3.353 (2.25, 4.44)	4.18E-06	TGM3	−4.555 (−6.13, −2.98)	3.10E-06
IFI6	2.554 (1.85, 3.25)	7.38E-06	ENDOU	−3.443 (−4.27, −2.61)	3.40E-06
GSDME	2.058 (1.65, 2.46)	3.88E-06	SPINK5	−3.595 (−4.56, −2.62)	2.93E-05
BST2	2.247 (1.67, 2.82)	2.89E-06	SPINK7	−3.577 (−4.82, −2.32)	4.90E-06

corresponding confidence intervals. The generated plot of the highest-ranked up- and down-regulated MD-DEGs common to both nodal positive and negative status are shown in Fig. 4.

3.4. Over-represented functions based on nodal status

ORA on the set of nodal negative OSCC DEGs identified through meta-analysis were associated with several biological processes, including responses to environmental stimuli like UV radiation, immune dysregulation such as interferon signalling, keratinocyte proliferation, and cellular responses to inflammatory cues like angiotensin were observed. These processes reflect potential mechanisms underlying the initiation of tumorigenesis in the initial stages (Supplementary Fig. 1 and Supplementary Table 5).

On the other hand, nodal positive DEGs revealed an enriched biological function related to wound healing, tissue remodelling, development & differentiation pathways, cell adhesion, migration, angiogenesis, and morphogenesis in the case of positive nodal status. These functions can be correlated to complex immune modulation processes, including positive regulation of leukocyte chemotaxis and regulation of inflammatory response, suggesting the tumour's ability to evade immune detection and create a favourable microenvironment (Supplementary Fig. 3 and Supplementary Table 7).

Alongside, discernible processes such as cellular functions and regulation, such as apoptotic signalling, protein maturation, response to lipopolysaccharide, and chemokine were enriched in both nodal positive and nodal negative statuses, suggesting their fundamental role in oral tumour biology and immune response.

KEGG over-representation analysis revealed 18 pathways in nodal negative and 15 in the nodal positive status of the DEG sets. Among these the most prominent, 'Immune related pathways' [IL-17 signalling pathway, Cytokine-cytokine receptor interaction, TNF signalling pathway]; Estrogen signalling pathway; Chemical carcinogenesis - DNA adducts and Infection-Related Pathways were observed in both conditions (Supplementary Figs. 2 and 4, Supplementary Tables 6 and 8).

3.5. PPI: hub genes underling OSCC progression

The PPI network constructed with input of MD-DEGs of nodal statuses revealed a network with 147 nodes and 1362 edges (Supplementary Fig. 5). The top three hub genes with the highest nodal degree (ND) in the specific subnetworks were: Nodal (−) PPI: CXCL1 (ND 46), GBP1 (ND 44), RTP4 (ND 42); Nodal (±)) PPI: CXCL10 (ND 76), MMP9 (ND 66), CXCL11 (ND 60); and Nodal (+) PPI: FN1 (ND 82), SERPINE1 (ND 50), IFI35 (ND 48) (Supplementary Table 4). The top three hub genes of each sub-network were considered to play a significant role in crosstalk across the early, intermediary, and advanced stages of the disease.

3.6. Potential genes underpinning the aggression nexus of OSCC

DMRs associated with OSCC recurrence were retrieved from the published literature by Sorroche et al. [20]. The DMRs meeting statistical False Discovery Rate <5 % were selected for intersection analysis with significant MD-DEGs from the current analysis of the OSCC-Nodal (+) group. The intersection revealed down-regulated genes CGNL1, MEIS1, GRHL3, and CFD.

The common genes were evaluated for potential prognostic clinical relevance through RFS and OS analysis using the Kaplan-Meier Plotter database. Among the genes analysed, only MEIS1 consistently exhibited a significant association, where low expression correlating with high risk of recurrence (HR = 0.45 and logrank P < 0.03) and worse OS outcomes compared to its high expression (Fig. 5 and Supplementary Fig. 6). Accordingly, MEIS1 was shortlisted as a candidate gene for further examination (Supplementary Table 1).

A total of 310 patient sample data of oral cavity sites in the GDC TCGA HNSC cohort were selected for validation processes using the UCSC Xena database (Fig. 6a). The Comparison of mRNA expression levels between normal and cancerous states revealed significant difference in the MEIS1 expression (down-regulated, p = 5.08E-06) (Fig. 6b)

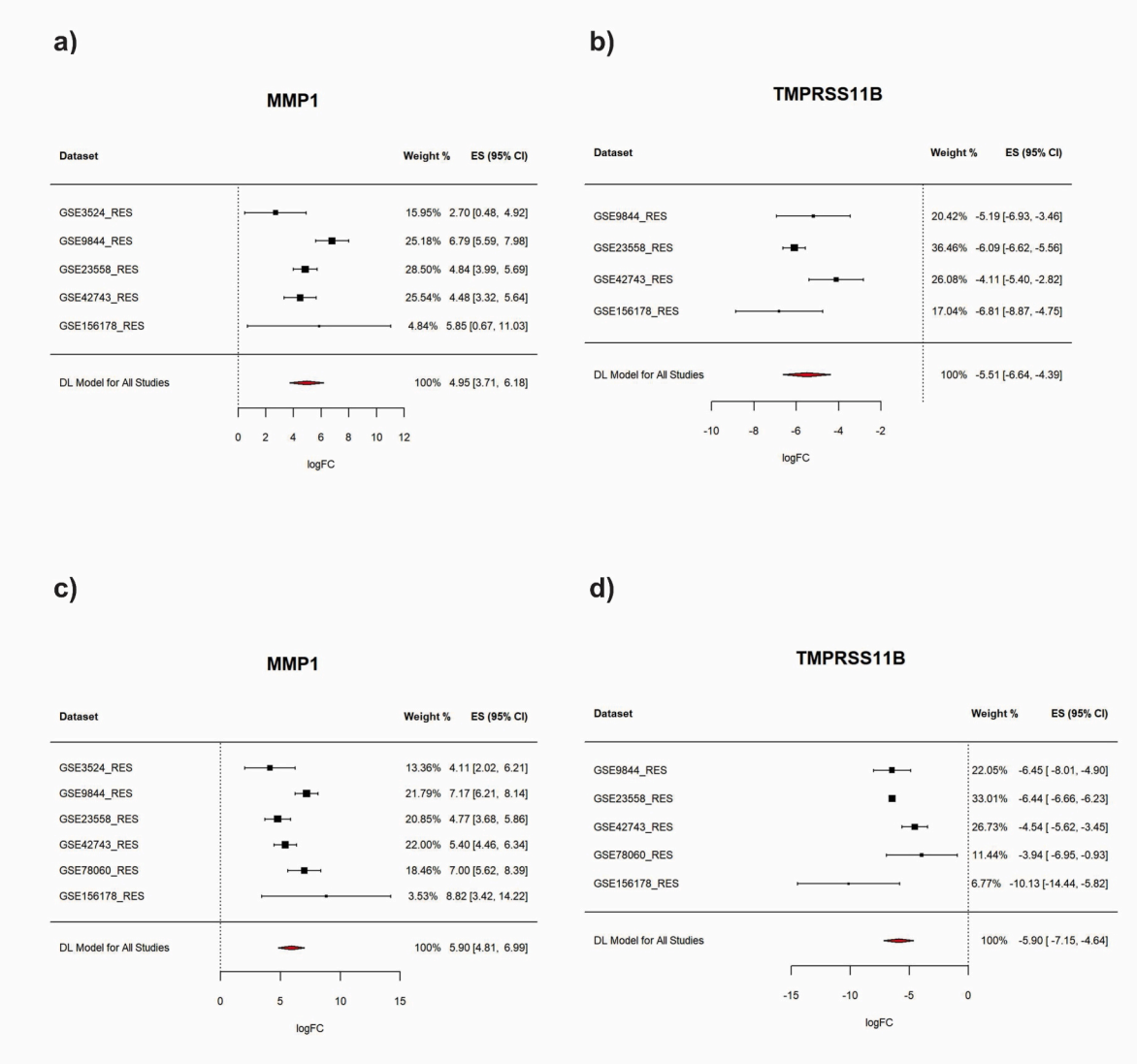

Fig. 4. Panels a and b illustrate forest plots for the highest-ranked upregulated (MMP1) and (TMPRSS11B) downregulated genes derived from the meta-analysis of nodal-negative OSCC DEGs. Similarly, panels c and d depict highest-ranked upregulated (MMP1) and downregulated (TMPRSS11B) genes derived from the meta-analysis of nodal-positive OSCC DEGs. Each plot illustrates the log2 fold changes for individual genes across specific datasets and the aggregated.

along with a notable difference in methylation status (hypermethylated) at CpG sites associated with OSCC recurrence (Fig. 6c). These observations on MEIS1 made from database were corroborated by its expression pattern (down-regulated, LFC -1.503, adj.p-val 4.40E-06) derived from the current OSCC-Nodal (+) meta-analysis and the methylation profile of the CpG sites associated with recurrence (hypermethylation, mean-betafc 0.08109) retrieved from the literature Sorroche et al. [20] (Supplementary Table 1). Furthermore, examination of the genomic locations of MEIS1 (NM_002398, TSS: 66435125) using the UCSC Genome Browser revealed a hypermethylation pattern in the upstream region, particularly at the LOC111258520 enhancer (NC_000002.12), which is located upstream of the promoter at the CpG sites: cg14775296 and cg06833110 (Fig. 6d).

3.7. Correlation between methylation patterns and mRNA expression of MEIS1

To examine the correlation between DNA methylation patterns and MEIS1 expression, we used the MEXPRESS database. Analysis of the TCGA-HNSC cohort, comprising 612 human samples, revealed a negative Pearson correlation between MEIS1 mRNA expression and the methylation status of CpG sites within the LOC111258520 enhancer region, as well as OSCC recurrence associated CpG sites (Supplementary Fig. 7).

4. Discussion

The meta-analysis of six publicly available transcriptomic profiles from multiple studies, encompassing nodal positive and nodal negative statuses in OSCC, was performed.

The integration of multiple datasets resulted in a substantial augmentation of the overall sample size, increasing it by 6.48 and 8.3 times from the median of individual OSCC nodal negative and positive dataset samples, respectively. This consolidative strategy enhanced the sample size leading to a more robust identification of significant up- or down-regulated DEGs within the meta-analysis framework, despite certain datasets displaying a minimal alteration in gene expression profiles. This resulted in the identification of 105 significant MD-DEGs in the nodal negative status and 171 significant MD-DEGs in the nodal positive status.

Furthermore, this comprehensive dataset amalgamation significantly diminished the risk of type-1 errors, which are often encountered

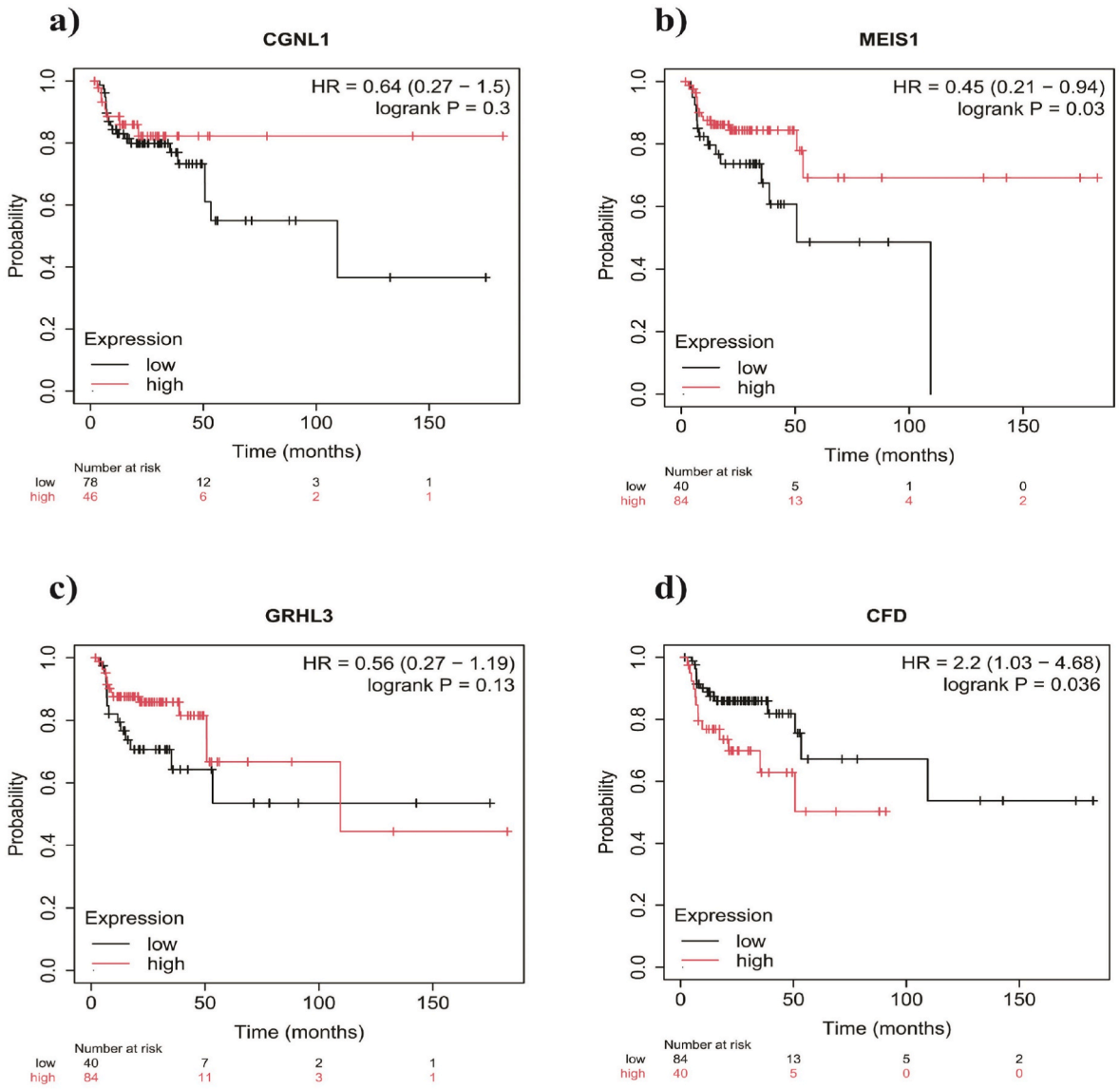

Fig. 5. Kaplan-Meier relapse free survival analysis of independent HNSC cohort: Assessment of a) CGNL1 b) MEIS1 c) GRHL3 d) CFD down-regulated OSCC lymph node metastatic genes with its potential association with recurrence.

when a gene is identified in a single dataset but not corroborated across others. The increased statistical power and reduction in false positives underscore the critical advantage of utilizing an aggregated strategy.

By adopting a series of analysis this study identified gene expression patterns involved in the disease progression. MD-DEGs were ranked using Topconfects, which prioritizes biologically significant genes by focusing on the confidence bounds of LFC, thus providing greater relevance compared to traditional p-value-based methods. The top 10 ranked up- and down-regulated genes were selected to delineate the molecular landscape of the disease. In the list, genes displaying significance in the nodal-negative state defined the early stage, while genes common to both nodal-negative and nodal-positive states marked the transition phase. Genes that are evident in the nodal-positive state were linked to the advanced stage. This framework was further supported by PPI analysis of MD-DEGs, which revealed critical hub genes involved in cellular crosstalk driving disease progression. The top three hub genes from each subnetwork were predicted to play key roles in early, transitional, and advanced stages. Additionally, ORA supported the biological differences between the nodal-positive and nodal-negative states (Fig. 7).

4.1. Early stage: molecular landscape of nodal negative status

A key event in the early phase of the disease is the upregulation of pro-inflammatory cytokines, such as CXCL1 (ND: 46), which significantly influences the interaction between tumour cells and the host microenvironment [45]. CXCL1 influences the growth of preneoplastic and malignant epithelial cells by facilitating the conversion of NOFs to CAFs through an autocrine mechanism. CAFs, characterized by their senescent fibroblasts, play a crucial role in promoting tumour progression [46]. Similarly, high ranked interferon-inducible cytokines in nodal negative status, such as CXCL10 (R-18) and CXCL11 (R-29) also contribute to progression of OSCC [47]. The initial phase of cancer development, marked by a surge in cytokines, transitions to a more advanced cancer state through the modulation of GBP1 (ND: 44) (guanylate binding protein 1) expression. The functional role of GBP1 in cancer is context-dependent on type of malignancy; however, our analysis reveals an elevated expression of GBP1, which is consistent with findings reported in HNSC [48].

Furthermore, the involvement of inflammasomes introduces additional complexity to the tumour microenvironment. We observed increased expression of AIM2 (R-20), which, in conjunction with IFI16

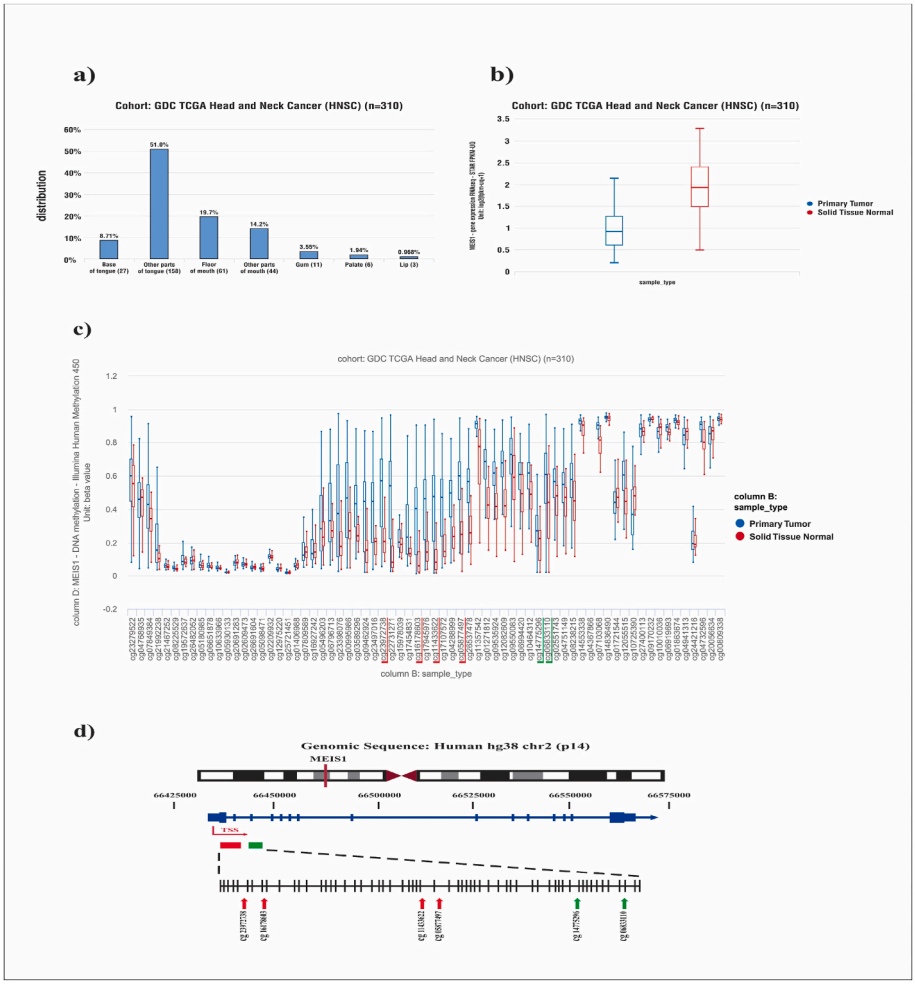

Fig. 6. Evaluation of the expression and methylation profiles of MEIS1 in OSCC: a) Selection of patient sample data related to oral cavity sites from GDC TCGA HNSC cohort in UCSC Xena database b) Comparison of the MEIS1 expression profile between normal tissue and OSCC patient samples c) Comparison of the MEIS1 methylation profile between normal and OSCC patient samples; red underscores denotes CpG sites associated with recurrence derived from literature Sorroche et al. [20], and green underscores indicates CpG sites associated with enhancer regulatory element d) Genomic locations of MEIS1 from the UCSC Genome Browser, including the chromosomal ideogram; the red line indicates the position of MEIS1 on chromosome 2, with the RefSeq gene map depicted by blue lines and bars; the red arrow denotes the transcription start site (TSS); recurrence-associated CpG regions are represented by red bar and enhancer regulatory element CpG regions are indicated by green bar.

(interferon gamma-inducible protein 16), forms a regulatory network that enhances malignant cell proliferation. This AIM2-IFI16 interaction facilitates the activation of the NF-κB signalling pathway, leading to the inhibition of apoptosis and thereby promoting tumour cell survival and progression [49]. The expression of PARP12 and RTDH were found to be dysregulated and are in alignment to previous study reports [50,51].

4.2. Intermediary stage: molecular landscape of the transitory phase to nodal positive status

As disease advances, established tumour cells coordinate diverse molecular pathways to promote a conducive microenvironment for an aggressive invasive front. The current analysis revealed that the majority of MD-DEGs of nodal negative status i.e., 92 out of 105 genes were observed to be transitioned into nodal metastatic status. Herein, the top-ranked genes inclusive of MMP1, PTHLH, ISG15, MMP12, and IF16; BST2 common between the nodal positive and negative status were considered to play an intermediatory role in the disease progression. Intriguingly, their expression change (LFC) increased during the transition from negative to positive status, indicating a potential correlation with cancer progression.

The intricate interplay of interferon signalling is prominently highlighted by the observed ectopic expressions of interferon-stimulated genes (ISGs) such as ISG15 and IF16 in the current analysis. ISG overexpression is positively correlated with enhanced expression of EGFR, initiating the processes of proliferation, invasion, and metastasis [52].

Further to add, the epithelial-mesenchymal transition (EMT) process signifies the transition to the metastatic stage by enhancing the mobility and invasiveness of the tumour cells. We observed altered gene expression related to extracellular matrix proteins (MMP) such as MMP-1, MMP-12, and MMP-9 within this intermediary category. The increased expression of MMPs leads to ECM degradation, initiating a significant alteration in cell phenotype as cells undergo EMT. Interestingly, our identification of MMP-9 in this context aligns, with its established role in promoting the formation of a pre-metastatic setting. MMP-9 also serves as a recognized marker of advanced disease stages [53].

The expression of other tumour modulators such as PTHLH and BST2 was also observed. Importantly, the regulation of the BST2 promoter is recognized to be mediated by the transcription factor STAT1. This STAT1/BST2 axis may modulate OSCC behaviour with its influence on AKT/ERK1/2 signalling pathway [54]. The abnormal expression of PTHLH has been reported to increase the key cell cycle regulators, including CCNA2, CCNE2, and CDC25A in HNSC [55].

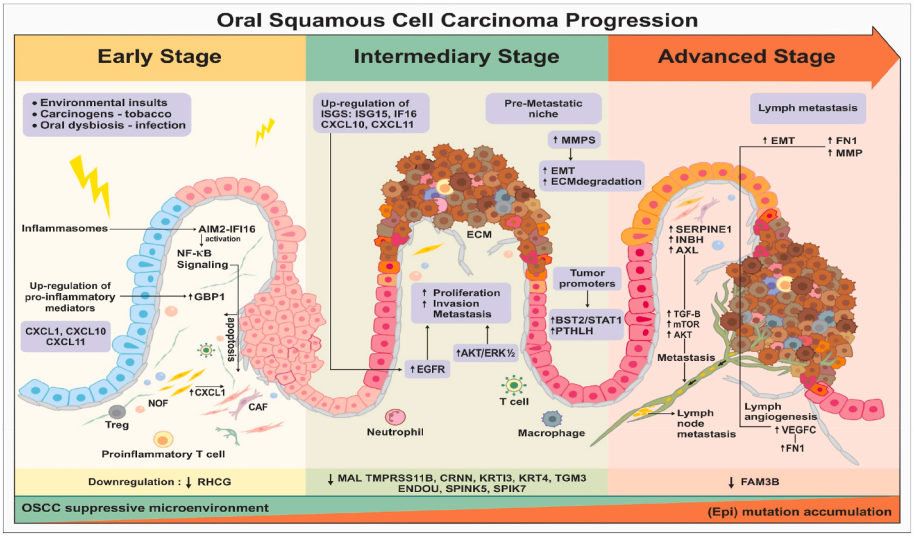

Fig. 7. Multistage progression of oral squamous cell carcinoma (OSCC). The progression of OSCC is characterized by a sequence of molecular and cellular changes within the oral epithelium, influenced by environmental factors like tobacco and infection. In the early stage, pro-inflammatory cytokines (e.g., CXCL1, CXCL10, CXCL11) are upregulated, driving the conversion of normal fibroblasts into cancer-associated fibroblasts (CAFs) and creating a tumour-promotive environment. As the disease progresses into the intermediary stage, interferon-stimulated genes (e.g., ISG15, IF16) and signalling pathways (e.g., EGFR, AKT/ERK1/2) are activated, facilitating tumour cell proliferation, invasion, and the establishment of a pre-metastatic niche through extracellular matrix degradation and EMT. In the advanced stage, further accumulation of genetic mutations and epigenetic alterations drives aggressive cancer behaviour, including lymph node metastasis, mediated by upregulated factors like SERPINE1, INHBA, and VEGFC. The tumour microenvironment evolves to support malignancy, characterized by dysfunctional immune responses and continuous remodelling of the ECM, which facilitates tumour seeding and metastasis. EMT epithelial-mesenchymal transition, CAF cancer-associated fibroblast, ECM extracellular matrix.

Taken together, these modifications with significantly down-regulated tumour suppressor genes develop a pre-metastatic niche that fosters an environment suitable for tumour seeding, and survival and expansion of metastatic cells.

4.3. Advanced stage: molecular landscape of nodal positive status

The commencement of the positive nodal metastatic phase is characterized by the migration of tumour cells. This migration is facilitated by the pre-metastatic niche, guided by the chemotactic gradient of chemokines within the tissue towards the sentinel lymph nodes [56]. The persistent involvement of MMP proteins throughout both the intermediary and advanced stages of the disease was evident in the current analysis. Herein, the highly ranked, upregulated genes MMP10 (R-5) and MMP3 (R-15) in nodal positive status profoundly impact the EMT process, thereby playing a key role in disease progression. Furthermore, the overexpression of MMP10 has been demonstrated to trigger the activation of AXL (receptor tyrosine kinase) downstream signalling molecules, including AKT, mTOR, and NF-κB. This upregulates the EMT process, resulting in increased cellular motility and subsequent lymph node metastasis [57].

The advanced stage of the disease was further characterized by the upregulation of several key proteins. Notably, the multifunctional cytokine inhibin subunit beta A (INHBA) (R-17), which is associated with TGF-β, exhibited increased expression. INHBA has been reported as a prognostic biomarker in HNSC, particularly in advanced stages [58].

Additionally, the upregulation of Interferon-Induced Protein 35 (IFI35) was observed. IFI35 (ND 48), known for its role in immune response modulation and regulation of interferon signalling, further underscores the complexity of the disease progression.

Accumulated genetic and epigenetic alteration in the tumour cells with disease progression can further dysregulate the pathways that form the hallmark of a mature tumour progressing towards metastasis. Another primary feature characterizing nodal positive status was the increased expression of fibronectin 1 (FN1). FN1 (ND 82) is a known promoter of the EMT process and is implicated in lymphangiogenesis, potentially through vascular endothelial growth factor-C (VEGF-C) [59]. Notably, we also observed a concurrent upregulation of VEGF-C expression in conjunction with FN1.

Gasdermin E (GSDME) (R-27), a key executor protein of pyroptosis, was found to be dysregulated in this study. Contrary to the findings reported by Wang et al. [60], we observed an upregulation of GSDME in with positive nodal status. Additionally, the expression intensity of the serine protease inhibitor SERPINE1(ND 50) was found to be increased. Elevated levels of SERPINE1 at the tumour-advancing front are known to correlate with greater disease aggressiveness [61]. These findings suggest a significant association between high SERPINE1 expression and the aggressive, nodal-positive nature of OSCC.

In sum, these traits serve as main conduits for lymph node metastasis.

Herein, the ranking analysis identified MMP1 as the highest-ranked upregulated gene, exhibiting a LFC of 4.946 (95 % CI: 3.71, 6.18) in nodal-negative samples, with a further increase to 5.899 (95 % CI: 4.80, 6.99) in nodal-positive samples, suggesting its persistent elevation across disease states. In parallel, TMPRSS11B was significantly down-regulated, with an LFC of −5.512 (95 % CI: −6.63, −4.38) in nodal-negative samples and −5.898 (95 % CI: −7.15, −4.64) in nodal-positive status. These findings highlight their potential as biomarkers, reinforcing their relevance in OSCC progression.

4.4. Findings on aggression nexus related to metastasis and recurrence of OSCC

Aberrant gene methylation has been acknowledged as a significant factor in cancer recurrence across diverse malignancies. This epigenetic modification can result in the silencing of tumour suppressor genes and the activation of oncogenes, thereby facilitating cancer progression and relapse [62].

To identify genes with altered expression and epigenetic regulation associated with lymph node metastasis and recurrence in OSCC, an intersection analysis was performed between current MD-DEG OSCC nodal (+) and the literature-derived DMRs associated with OSCC

recurrence.

This revealed a panel of DEGs with aberrant methylation profile including CGNL1, MEIS1, GRHL3 and CFD. Following validation and assessments in external dataset, MEIS1 was identified as a candidate nodal metastatic gene characterized by hypo-expression. Examination of the methylation status in the TCGA-HNSC cohort revealed a significant hypermethylation pattern, particularly in the upstream promoter regions of the gene. Notably, an inverse correlation was observed between this hypermethylation and gene expression at the regulatory enhancer region. Moreover, a significant association of lower MEIS1 expression with unfavourable RFS and OS, in conjunction with comparable hypermethylation of CpG sites associated with OSCC recurrence state, indicates a complex role of MEIS1 in both metastatic processes and the recurrence of OSCC.

MEIS1 is a member of TALE (3-amino-acid loop extension) family of homeodomain transcription factors (TF), which interacts with Hox transcription factors involved in regulating developmental processes and maintaining tissue differentiation [63]. MEIS1 encompasses a large genomic region, with the majority consisting of non-coding sequences. This structure accommodates the possibility of multiple interacting *cis*-regulatory elements that govern its expression. The findings of Xiang et al. demonstrate the importance of regulatory roles of both the promoter and enhancer regions in MEIS1 transcription in Acute Myeloid Leukemia (AML). The authors demonstrated that the loss of specific enhancer regions, particularly within intron 6, resulted in a marked reduction in MEIS1 expression. Furthermore, FLI1, an ETS family transcription factor, was shown to facilitate chromatin looping between enhancers. Additionally, a stronger interaction between the distant downstream enhancers E2 or E3 and the promoter region was shown to increase MEIS1 transcription [64,65].

Any aberrant expression of MEIS1 has been implicated in various cancers and is context dependent. The present analysis reveals the hypo-expression and hypermethylation patterns of MEIS1 in OSCC, which is in similar lines with findings in other malignancies, including renal cell carcinoma and non-small-cell lung cancer [66,67]. Moreover, in prostate cancer, epigenetic silencing through promoter hypermethylation has been associated with a stepwise reduction in MEIS1 expression, from benign tissue to localized disease, with an additional decline observed in metastatic conditions [68]. Moreover, its abnormal expression is associated with transcriptional misregulation, which facilitates tumorigenesis and enhances tumour aggressiveness [69]. In contrast, studies on AML have shown that mutations in DNA methyltransferase 3A (DNMT3A) lead to MEIS1 hypomethylation, which drives its overexpression and promotes leukemogenesis [70]. Overall, these findings underscore the intricate epigenetic regulation of MEIS1 and its relevance in tumor progression.

Metastasis and recurrence are two facets of a common nexus that drive the aggressiveness of the cancer. The evolutionary trajectory of cancer cells is facilitated by various factors, including genetic mutations, dysregulated pathways, and the tumor microenvironment. Among these, transcription factors (TFs) play a vital role in cell reprogramming, including EMT, through their interaction with various signalling pathways. This reprogramming not only enhances tumour plasticity and invasiveness but also sustains cancer stem cell (CSC) properties. By preserving pluripotency and self-renewal, it supports metastasis-initiating cells and fosters an invasive niche. The slow-proliferating nature and dormant behaviour of CSCs enable them to withstand anti-cancer therapy and are thought to give rise to minimal residual disease (MRD). These cells can persist in an undetectable state for years or even decades, eventually reactivating under favourable conditions to drive tumour recurrence [71–73].

To the best of our knowledge, this transcriptomic meta-analysis with cumulative evidence presents the first investigation into the nexus of aggression involving recurrence and metastasis in OSCC. The analysis demonstrates the downregulated and aberrant methylation patterns of MEIS1 a homeobox TF, in lymph node metastasis and disease recurrence, indicating its involvement in promoting OSCC aggressiveness. It is evident that MEIS1 silencing suppresses key CDK inhibitors, Cdkn1a (p21) and Arf (p14ARF), in lung cancer cell lines, potentially leading to increased CDK activity. This dysregulation of cell cycle regulators, known for their role in modulating cancer stemness, may contribute progression of the disease [74]. Given its significance in OSCC, MEIS1 has substantial potential as a prognostic and therapeutic biomarker, meriting further investigation and attention. Homeobox genes, known for their role in tumorigenesis, have been widely recognized as promising biomarkers for cancer diagnosis, prognosis, and therapy such as bladder cancer. In this context, studies on MEIS1 as a methylation marker and expression signature, alongside other homeobox genes, have demonstrated predictive models with accuracies of up to 97 % in various malignancies. Intriguingly, its expression has also been linked to immune checkpoints such as PD-L1 and PD-L2, suggesting a potential role in predicting therapy responses [75].

5. Conclusion

In summary, this study presents a comprehensive meta-analysis of independent gene expression datasets, offering valuable insights into disease progression and highlighting the reliability of genes associated with lymph node status. The study identifies MEIS1 within the nexus of OSCC aggressiveness as a shared determinant of lymph node metastasis and recurrence. While these findings highlight its prognostic relevance, further functional studies by in-vitro and in-vivo experiments are warranted to confirm its role in imparting OSCC aggressiveness.

CRediT authorship contribution statement

Soujanya J. Vastrad: Writing – original draft, Visualization, Formal analysis, Data curation, Conceptualization. **Ganesan Rajalekshmi Saraswathy:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Jagadish B. Dasari:** Writing – review & editing, Validation, Data curation, Conceptualization. **Gouri Nair:** Writing – review & editing, Validation, Data curation, Conceptualization. **Ashok Madarakhandi:** Writing – review & editing, Validation, Data curation, Conceptualization. **Dominic Augustine:** Writing – review & editing, Conceptualization. **S.V. Sowmya:** Writing – review & editing, Conceptualization.

Data availability

The transcriptomic datasets analysed in this study are accessible from the NCBI Gene Expression Omnibus database. Data related to differentially methylated regions associated with OSCC recurrence were obtained with the authors' permission from the published work of Sorroche et al.

Funding

No funding was received.

Declaration of competing interest

The authors have no competing interests to declare.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbrep.2025.102001.

References

[1] Y. Tan, Z. Wang, M. Xu, et al., Oral squamous cell carcinomas: state of the field and emerging directions, Int. J. Oral Sci. 15 (1) (2023) 44, https://doi.org/10.1038/s41368-023-00249-w.

[2] National Cancer Institute, Cancer statistics explorer network. https://seer.cancer.gov/statistics-network/explorer/. (Accessed 20 February 2024).

[3] J.A. Woolgar, A. Triantafyllou, J.S. Lewis Jr., et al., Prognostic biological features in neck dissection specimens, Eur. Arch. Otorhinolaryngol. 270 (5) (2013) 1581–1592, https://doi.org/10.1007/s00405-012-2170-9.

[4] S.H. Huang, B. O'Sullivan, Oral cancer: current role of radiotherapy and chemotherapy, Med. Oral Patol. Oral Cir. Bucal 18 (2) (2013) e233–e240, https://doi.org/10.4317/medoral.18772.

[5] G. Studer, R.A. Zwahlen, K.W. Graetz, B.J. Davis, et al., IMRT in oral cavity cancer, Radiat. Oncol. 2 (2007) 16, https://doi.org/10.1186/1748-717x-2-16.

[6] V. Espeli, A. Gomez, P. Balermpas, et al., A systematic review of contemporary phase I trials in patients with recurrent/metastatic head and neck cancer, Oral Oncol. Rep. 11 (2024) 100588, https://doi.org/10.1016/j.oor.2024.100588.

[7] M. Partridge, G. Emilion, S. Pateromichelakis, et al., The prognostic significance of allelic imbalance at key chromosomal loci in oral cancer, Br. J. Cancer 79 (11–12) (1999) 1821–1827, https://doi.org/10.1038/sj.bjc.6690290.

[8] F. Beegum, A. Trimukhe, Traversing the terrain: potential pitfalls within the AJCC 8th edition staging system for lip and oral cavity cancers, Head Neck Pathol 18 (1) (2024) 62, https://doi.org/10.1007/s12105-024-01663-0.

[9] Y.R. Murciano-Goroff, B.S. Taylor, D.M. Hyman, et al., Toward a more precise future for oncology, Cancer Cell 37 (4) (2020) 431–442, https://doi.org/10.1016/j.ccell.2020.03.014.

[10] J. Parsons, C. Francavilla, 'Omics approaches to explore the breast cancer landscape, Front. Cell Dev. Biol. 7 (2020) 395, https://doi.org/10.3389/fcell.2019.00395.

[11] P. de Valpine, H.M. Bitter, M.P. Brown, et al., A simulation-approximation approach to sample size planning for high-dimensional classification studies, Biostatistics 10 (3) (2009) 424–435, https://doi.org/10.1093/biostatistics/kxp001.

[12] B. Linggi, V. Jairath, G. Zou, et al., Meta-analysis of gene expression disease signatures in colonic biopsy tissue from patients with ulcerative colitis, Sci. Rep. 11 (1) (2021) 18243, https://doi.org/10.1038/s41598-021-97366-5.

[13] A. López-Cerdán, Z. Andreu, M.R. Hidalgo, et al., Unveiling sex-based differences in Parkinson's disease: a comprehensive meta-analysis of transcriptomic studies, Biol. Sex Differ. 13 (1) (2022) 68, https://doi.org/10.1186/s13293-022-00477-5.

[14] V.R. Umapathy, P.M. Natarajan, B. Swamikannu, Comprehensive review on development of early diagnostics on oral cancer with a special focus on biomarkers, Appl. Sci. 12 (10) (2022) 4926, https://doi.org/10.3390/app12104926.

[15] M. Abedi, R. Fatehi, K. Moradzadeh, et al., Big data to knowledge: common pitfalls in transcriptomics data analysis and representation, RNA Biol. 16 (11) (2019) 1531–1533, https://doi.org/10.1080/15476286.2019.1652525.

[16] Y. Sun, Z. Sang, Q. Jiang, et al., Transcriptomic characterization of differential gene expression in oral squamous cell carcinoma: a meta-analysis of publicly available microarray data sets, Tumor Biol. (2016), https://doi.org/10.1007/s13277-016-5439-6.

[17] W. Lu, Y. Wang, M. Gan, et al., Prognosis and predictive value of heat-shock proteins expression in oral cancer: a PRISMA-compliant meta-analysis, Medicine (Baltim.) 100 (3) (2021) e24274, https://doi.org/10.1097/md.0000000000024274.

[18] S.L. Luo, Y.G. Xie, Z. Li, et al., E-cadherin expression and prognosis of oral cancer: a meta-analysis, Tumor Biol. 35 (6) (2014) 5533–5537, https://doi.org/10.1007/s13277-014-1728-0.

[19] S. Shaikh, D.K. Yadav, K. Bhadresha, et al., Integrated computational screening and liquid biopsy approach to uncover the role of biomarkers for oral cancer lymph node metastasis, Sci. Rep. 13 (1) (2023) 14033, https://doi.org/10.1038/s41598-023-41348-2.

[20] B.P. Sorroche, F.R. Talukdar, S.C.S. Lima, et al., DNA methylation markers from negative surgical margins can predict recurrence of oral squamous cell carcinoma, Cancers 13 (12) (2021) 2915, https://doi.org/10.3390/cancers13122915.

[21] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2023. https://www.R-project.org.

[22] T. Barrett, S.E. Wilhite, P. Ledoux, et al., NCBI GEO: archive for functional genomics data sets–update, Nucleic Acids Res. 41 (Database issue) (2013) D991–D995.

[23] D. Moher, A. Liberati, J. Tetzlaff, et al., Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement, PLoS Med. 6 (7) (2009) e1000097, https://doi.org/10.1371/journal.pmed.1000097.

[24] S. Davis, P.S. Meltzer, GEOquery: a bridge between the gene expression Omnibus (GEO) and BioConductor, Bioinformatics 23 (14) (2007) 1846–1847, https://doi.org/10.1093/bioinformatics/btm254.

[25] S. Durinck, P.T. Spellman, E. Birney, et al., Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt, Nat. Protoc. 4 (8) (2009) 1184–1191, https://doi.org/10.1038/nprot.2009.97.

[26] S. Tweedie, B. Braschi, K. Gray, et al., Genenames.org: the HGNC and VGNC resources in 2021, Nucleic Acids Res. 49 (D1) (2021) D939–D946, https://doi.org/10.1093/nar/gkaa980.

[27] M.E. Ritchie, B. Phipson, D. Wu, et al., Limma powers differential expression analyses for RNA-sequencing and microarray studies, Nucleic Acids Res. 43 (7) (2015) e47, https://doi.org/10.1093/nar/gkv007, 2015.

[28] K. Blighe, S. Rana, M. Lewis, Enhanced Volcano: publication-ready volcano plots with enhanced colouring and labelling, R package version 1.13.2, https://github.com/kevinblighe/EnhancedVolcano, 2024.

[29] A. Kassambara, F. Mundt, Factoextra: extract and visualize the results of multivariate data analyses: R package version 1.0.7. https://CRAN.R-project.org/package=factoextra, 2020.

[30] W. Viechtbauer, Conducting meta-analyses in R with the metafor package, J. Stat. Software 36 (3) (2010) 1–48, https://doi.org/10.18637/jss.v036.i03.

[31] R. DerSimonian, N. Laird, Meta-analysis in clinical trials, Control. Clin. Trials 7 (3) (1986) 177–188, https://doi.org/10.1016/0197-2456(86)90046-2.

[32] A. Ludt, A. Ustjanzew, H. Binder, et al., Interactive and reproducible workflows for exploring and modeling RNA-seq data with pcaExplorer, ideal, and GeneTonic, Curr. Protoc. 2 (4) (2022) e411, https://doi.org/10.1002/cpz1.411.

[33] P.F. Harrison, A.D. Pattison, D.R. Powell, et al., Topconfects: a package for confident effect sizes in differential expression analysis provides a more biologically useful ranked gene list, Genome Biol. 20 (1) (2019) 67, https://doi.org/10.1186/s13059-019-1674-7.

[34] T. Wu, E. Hu, S. Xu, et al., clusterProfiler 4.0: a universal enrichment tool for interpreting omics data, Innovation 2 (3) (2021) 100141, https://doi.org/10.1016/j.xinn.2021.100141.

[35] M. Ashburner, C.A. Ball, J.A. Blake, et al., Gene ontology: tool for the unification of biology. The Gene Ontology Consortium, Nat. Genet. 25 (1) (2000) 25–29, https://doi.org/10.1038/75556.

[36] M. Kanehisa, S. Goto, KEGG: kyoto encyclopedia of genes and genomes, Nucleic Acids Res. 28 (1) (2000) 27–30, https://doi.org/10.1093/nar/28.1.27.

[37] D. Szklarczyk, A.L. Gable, K.C. Nastou, et al., The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets, Nucleic Acids Res. 49 (2021) D605–D612, https://doi.org/10.1093/nar/gkaa1074.

[38] P. Shannon, A. Markiel, O. Ozier, et al., Cytoscape: a software environment for integrated models of biomolecular interaction networks, Genome Res. 13 (11) (2003) 2498–2504, https://doi.org/10.1101/gr.1239303.

[39] B. Győrffy, Integrated analysis of public datasets for the discovery and validation of survival-associated genes in solid tumors, Innovation 5 (3) (2024) 100625, https://doi.org/10.1016/j.xinn.2024.100625.

[40] M.J. Goldman, B. Craft, M. Hastie, et al., Visualizing and interpreting cancer genomics data via the Xena platform, Nat. Biotechnol. 38 (6) (2020) 675–678, https://doi.org/10.1038/s41587-020-0546-8.

[41] A. Koch, T. De Meyer, J. Jeschke, et al., MEXPRESS: visualizing expression, DNA methylation and clinical TCGA data, BMC Genom. 16 (1) (2015) 636, https://doi.org/10.1186/s12864-015-1847-z.

[42] A. Koch, J. Jeschke, W. Van Criekinge, et al., MEXPRESS update 2019, Nucleic Acids Res. 47 (W1) (2019) W561–W565, https://doi.org/10.1093/nar/gkz445.

[43] G. Perez, G.P. Barber, A. Benet-Pages, et al., The UCSC Genome Browser database: 2025 update, Nucleic Acids Res. (2024), https://doi.org/10.1093/nar/gkae974.

[44] A. Suzuki, S. Kawano, T. Mitsuyama, et al., DBTSS/DBKERO for integrated analysis of transcriptional regulation, Nucleic Acids Res. 46 (D1) (2018) D229–D238, https://doi.org/10.1093/nar/gkx1001.

[45] L.Y. Wei, J.J. Lee, C.Y. Yeh, et al., Reciprocal activation of cancer-associated fibroblasts and oral squamous carcinoma cells through CXCL1, Oral Oncol. 88 (2019) 115–123, https://doi.org/10.1016/j.oraloncology.2018.11.002.

[46] E.K. Kim, S. Moon, D.K. Kim, et al., CXCL1 induces senescence of cancer-associated fibroblasts via autocrine loops in oral squamous cell carcinoma, PLoS One 13 (1) (2018) e0188847, https://doi.org/10.1371/journal.pone.0188847.

[47] F. Geng, Q. Wang, C. Li, et al., Identification of potential candidate genes of oral cancer in response to chronic infection with porphyromonas gingivalis using bioinformatical analyses, Front. Oncol. 9 (2019) 91, https://doi.org/10.3389/fonc.2019.00091.

[48] A.T. Honkala, D. Tailor, S.V. Malhotra, Guanylate-binding protein 1: an emerging target in inflammation and cancer, Front. Immunol. 10 (2020) 3139, https://doi.org/10.3389/fimmu.2019.03139.

[49] L. Wang, L. Sun, K.M. Byrd, et al., AIM2 inflammasome's first decade of discovery: focus on oral diseases, Front. Immunol. 11 (2020) 1487, https://doi.org/10.3389/fimmu.2020.01487.

[50] D. Das, A. Maitra, C.K. Panda, et al., Genes and pathways monotonically dysregulated during progression from normal through leukoplakia to gingivo-buccal oral cancer, NPJ Genom Med 6 (1) (2021) 32, https://doi.org/10.1038/s41525-021-00195-8.

[51] A. Khammanivong, A. Anandharaj, X. Qian, et al., Transcriptome profiling in oral cavity and esophagus tissues from (S)-N'-nitrosonornicotine-treated rats reveals candidate genes involved in human oral cavity and esophageal carcinogenesis, Mol. Carcinog. 55 (12) (2016) 2168–2182, https://doi.org/10.1002/mc.22459.

[52] Y. Yuan, H. Qin, H. Li, et al., The functional roles of ISG15/ISGylation in cancer, Molecules 28 (3) (2023) 1337, https://doi.org/10.3390/molecules28031337.

[53] M. Aparna, L. Rao, V. Kunhikatta, et al., The role of MMP-2 and MMP-9 as prognostic markers in the early stages of tongue squamous cell carcinoma, J. Oral Pathol. Med. 44 (5) (2015) 345–352, https://doi.org/10.1111/jop.12245.

[54] F. Shan, S. Shen, X. Wang, et al., BST2 regulated by the transcription factor STAT1 can promote metastasis, invasion and proliferation of oral squamous cell carcinoma via the AKT/ERK1/2 signaling pathway, Int. J. Oncol. 62 (4) (2023) 54, https://doi.org/10.3892/ijo.2023.5502.

[55] W.M. Chang, Y.F. Lin, C.Y. Su, et al., Parathyroid hormone-like hormone is a poor prognosis marker of head and neck cancer and promotes cell growth via RUNX2 regulation, Sci. Rep. 7 (2017) 41131, https://doi.org/10.1038/srep41131.

[56] L.C. Dieterich, C. Tacconi, L. Ducoli, et al., Lymphatic vessels in cancer, Phys. Rev. (4) (2022) 1837–1879, https://doi.org/10.1152/physrev.00039.2021.

[57] B. Dharavath, A. Butle, A. Pal, et al., Role of miR-944/MMP10/AXL- axis in lymph node metastasis in tongue cancer, Commun. Biol. 6 (1) (2023) 57, https://doi.org/10.1038/s42003-023-04437-6.
[58] S. Zhang, K. Jin, T. Li, et al., Comprehensive analysis of INHBA: a biomarker for anti-TGFβ treatment in head and neck cancer, Exp Biol Med (Maywood) 247 (15) (2022) 1317–1329, https://doi.org/10.1177/15353702221085203.
[59] Y. Morita, K. Hata, M. Nakanishi, et al., Cellular fibronectin 1 promotes VEGF-C expression, lymphangiogenesis and lymph node metastasis associated with human oral squamous cell carcinoma, Clin. Exp. Metastasis 32 (7) (2015) 739–753, https://doi.org/10.1007/s10585-015-9741-2.
[60] S. Wang, M.J. Zhang, Z.Z. Wu, et al., GSDME is related to prognosis and response to chemotherapy in oral cancer, J. Dent. Res. 101 (7) (2022) 848–858, https://doi.org/10.1177/00220345211073072.
[61] J. Dhanda, A. Triantafyllou, T. Liloglou, et al., SERPINE1 and SMA expression at the invasive front predict extracapsular spread and survival in oral squamous cell carcinoma, Br. J. Cancer 111 (11) (2014) 2114–2121, https://doi.org/10.1038/bjc.2014.500.
[62] S.B. Baylin, P.A. Jones, A decade of exploring the cancer epigenome - biological and translational implications, Nat. Rev. Cancer 11 (10) (2011) 726–734, https://doi.org/10.1038/nrc3130.
[63] D.M. Wellik, M. Torres, M.A. Ros, Forward to the special issue on Hox/Tale transcription factors in development and disease, Dev. Dyn. 243 (1) (2014) 1–3, https://doi.org/10.1002/dvdy.24098.
[64] P. Xiang, X. Yang, L. Escano, et al., Elucidating the importance and regulation of key enhancers for human MEIS1 expression, Leukemia 36 (8) (2022) 1980–1989, https://doi.org/10.1038/s41375-022-01602-4.
[65] P. Xiang, W. Wei, C. Lo, et al., Delineating MEIS1 cis-regulatory elements active in hematopoietic cells, Leukemia 28 (2) (2013) 433–436, https://doi.org/10.1038/leu.2013.287.
[66] J. Zhu, L. Cui, A. Xu, et al., MEIS1 inhibits clear cell renal cell carcinoma cells proliferation and in vitro invasion or migration, BMC Cancer 17 (1) (2017) 176, https://doi.org/10.1186/s12885-017-3155-2.
[67] W. Li, K. Huang, H. Guo, et al., Meis1 regulates proliferation of non-small-cell lung cancer cells, J. Thorac. Dis. 6 (6) (2014) 850–855, https://doi.org/10.3978/j.issn.2072-1439.2014.06.03.
[68] R.R. Bhanvadia, C. VanOpstall, H. Brechka, et al., MEIS1 and MEIS2 expression and prostate cancer progression: a role for HOXB13 binding partners in metastatic disease, Clin. Cancer Res. 24 (15) (2018) 3668–3680, https://doi.org/10.1158/1078-0432.ccr-17-3673.
[69] H. Li, Y. Tang, L. Hua, et al., A systematic pan-cancer analysis of MEIS1 in human tumors as prognostic biomarker and immunotherapy target, J. Clin. Med. 12 (4) (2023) 1646, https://doi.org/10.3390/jcm12041646.
[70] H.J. Ferreira, H. Heyn, M. Vizoso, et al., DNMT3A mutations mediate the epigenetic reactivation of the leukemogenic factor MEIS1 in acute myeloid leukemia, Oncogene 36 (29) (2017) 4233, https://doi.org/10.1038/onc.2017.57.
[71] J. Majidpoor, K. Mortezaee, Steps in metastasis: an updated review, Med. Oncol. 38 (1) (2021) 3, https://doi.org/10.1007/s12032-020-01447-w. Published 2021 Jan 4.
[72] C. Peitzsch, A. Tyutyunnykova, K. Pantel, A. Dubrovska, Cancer stem cells: the root of tumor recurrence and metastases, Semin. Cancer Biol. 44 (2017) 10–24, https://doi.org/10.1016/j.semcancer.2017.02.011.
[73] H.Y. Min, H.Y. Lee, Cellular dormancy in cancer: mechanisms and potential targeting strategies, Cancer Res Treat 55 (3) (2023) 720–736, https://doi.org/10.4143/crt.2023.468.
[74] W. Li, K. Huang, H. Guo, et al., Meis1 regulates proliferation of non-small-cell lung cancer cells, J. Thorac. Dis. 6 (6) (2014) 850–855, https://doi.org/10.3978/j.issn.2072-1439.2014.06.03.
[75] F.W. Chin, S.C. Chan, A. Veerakumarasivam, Homeobox gene expression dysregulation as potential diagnostic and prognostic biomarkers in bladder cancer, Diagnostics 13 (16) (2023) 2641, https://doi.org/10.3390/diagnostics13162641.