Original Article

Identification of an Exploratory Three-Gene mRNA Model for Mucoepidermoid Carcinoma Using Public RNA-Seq Data

Kiarash Fereidouni [1], DMD; Shahram Hamedani [2], MScD;

[1] Orthodontic Research Center, School of Dentistry, Shiraz University of Medical Sciences, Shiraz, Iran.
[2] Oral and Dental Disease Research Center, School of Dentistry, Shiraz University of Medical Sciences, Shiraz, Iran.

KEY WORDS
Mucoepidermoid carcinoma;
Parotid gland;
RNA sequencing;
Gene-expression signature;
Tumor biomarkers;

ABSTRACT

Background: Mucoepidermoid carcinoma (MEC) is the most common malignant salivary gland tumor; however, comprehensive molecular models that distinguish MEC from normal parotid tissue remain poorly defined.

Purpose: This study aimed to identify and internally evaluate a compact candidate expression signature distinguishing MEC from normal parotid tissue using a single public RNA-sequencing cohort.

Materials and Method: In this retrospective bioinformatics/in silico observational study, RNA-sequencing data from GSE282430, including 21 MEC and 6 normal parotid samples, were analyzed using differential expression analysis and Hallmark pathway enrichment. Recently published MEC-related gene signatures were reconstructed, harmonized, and intersected with the cohort. Logistic regression and least absolute shrinkage and selection operator (LASSO) regression were then used to evaluate tumor-versus-normal discrimination and derive a compact panel.

Results: A total of 385 genes were differentially expressed between MEC and normal tissue. A literature-derived union panel of 27 genes achieved an area under the receiver operating characteristic curve (AUC) of 1.00 for distinguishing MEC from normal parotid tissue. LASSO-based reduction yielded an exploratory three-gene signature comprising MLXIPL, POSTN, and SPP1, which showed an exploratory cross-validated AUC of 0.976. The selected genes represent complementary metabolic, stromal/extracellular matrix, and immune matrix processes, although functional interactions among them were not experimentally assessed.

Conclusion: Integration of four literature-derived gene sets identified an exploratory MLXIPL-POSTN-SPP1 candidate expression signature that distinguished MEC from normal parotid tissue within a small public RNA-sequencing cohort. These findings support further investigation of this candidate signature, but external cross-platform validation and experimental confirmation are required before diagnostic or clinical interpretation.

Received: 13 May 2026;
Revised: 28 June 2026;
Accepted: 6 September 2026;

Corresponding Author: Fereidouni K, School of Dentistry, Shiraz University of Medical Sciences, Ghasroddasht Street, Ghomabad Ave, Shiraz, Iran. Tel: +989177859489 Email: kiarashfereidouni@gmail.com

Cite this article as:

Introduction

Mucoepidermoid carcinoma (MEC) is the most common malignant salivary gland tumor and poses diagnostic and therapeutic challenges due to its histological heterogeneity and variable clinical behavior [1]. Management primarily involves surgery, with adjuvant radiotherapy or chemotherapy considered based on tumor grade, stage, resection status, and other clinicopathological risk factors [2-3]. Although machine learning has been explored for outcome prediction in salivary gland

malignancies, molecular tools for MEC remain limited [14].

The biological complexity of MEC, including its mixed composition of mucous, epidermoid, and intermediate cells, contributes to diagnostic and grading challenges [5-7]. Recent research on the tumor microenvironment and immune landscape has advanced understanding of salivary gland tumor biology and identified candidate molecular biomarkers [8-10]. However, because MEC is rare, its molecular features remain less well characterized than those of more common solid tumors.

Previous transcriptomic studies have identified MEC-related gene sets with potential biological and discriminatory relevance [11-16]. However, whether these published genes retain tumor-versus-normal discriminatory information in another transcriptomic cohort and whether they can be reduced to a compact signature remain uncertain. This study therefore evaluated four literature-derived MEC gene sets in GSE282430 and explored whether their union could yield a parsimonious candidate expression signature to distinguish MEC from normal parotid tissue.

Materials and Method

In this retrospective observational bioinformatics study, Bulk RNA-sequencing data were obtained from the Gene Expression Omnibus (GEO) dataset GSE282430. Samples annotated as human tumor ("hT") or human normal ("hN") parotid tissue were retained, yielding 21 MEC and 6 normal samples. Normalized gene-expression data were imported into R version 4.5.1. Rows without valid gene symbols were removed, duplicate symbols were collapsed by their median expression, and protein-coding, immunoglobulin constant-region, and T-cell receptor constant-region genes with non-zero variance were retained.

Differential expression between MEC and normal parotid tissue was assessed using Limma with empirical Bayes moderation. The treat function was applied with lfc = 1, thereby testing differential expression relative to a minimum absolute $\log_2$ fold-change threshold of 1. Genes with FDR < 0.05 were considered significantly differentially expressed.

Pre-ranked Hallmark enrichment analysis was performed using fgsea, with moderated t-statistics as the ranking metric. Hallmark gene sets were obtained from MSigDB, and pathways with FDR q < 0.05 were considered significantly enriched.

Four MEC-related gene signatures published after 2020 were reconstructed from the literature, harmonized to official HUGO Gene Nomenclature Committee (HGNC) symbols, and intersected with GSE282430. These included signatures from Naakka *et al.* [13], Kang *et al.* [14], Ramírez-Martínez *et al.* [15], and Zou *et al.* [16]. The intersected genes were evaluated as individual panels and as a 27-gene union panel.

MEC versus normal parotid status was modeled using L1-penalized logistic regression. Internal evaluation was performed using five-fold nested cross-validation, with an outer stratified loop for held-out predictions and an inner cross-validation loop for selection of the regularization parameter using the lambda.1se rule. Genes selected with non-zero coefficients in at least 60% of outer folds defined the compact candidate signature. The area under the receiver operating characteristic curve (AUC), 95% confidence interval (CI), and Brier score were calculated from aggregated held-out predictions. These analyses represent internal cross-validation and not external model validation. No probability threshold was proposed because the sample size and absence of external validation did not support estimation of a clinically transportable decision cutoff. Histologic grade was not included as a predictor because the prespecified objective was binary discrimination between MEC and normal parotid tissue rather than grade classification.

Results

Twenty-one MEC samples and six normal parotid samples from GSE282430 were analyzed. Principal component analysis showed clear separation between MEC and normal samples, with minimal overlap (Figure 1a). Differential expression analysis identified 385 genes, including 130 upregulated and 255 downregulated in MEC. The volcano plot highlighted markedly dysregulated genes (Figure 1b), while the heatmap showed clustering of samples by diagnosis (Figure 1c).

Hallmark enrichment analysis showed positive enrichment for epithelial-mesenchymal transition, extra cellular matrix-related, cell cycle, and immune/ inflammatory pathways in MEC. In contrast, metabolic and

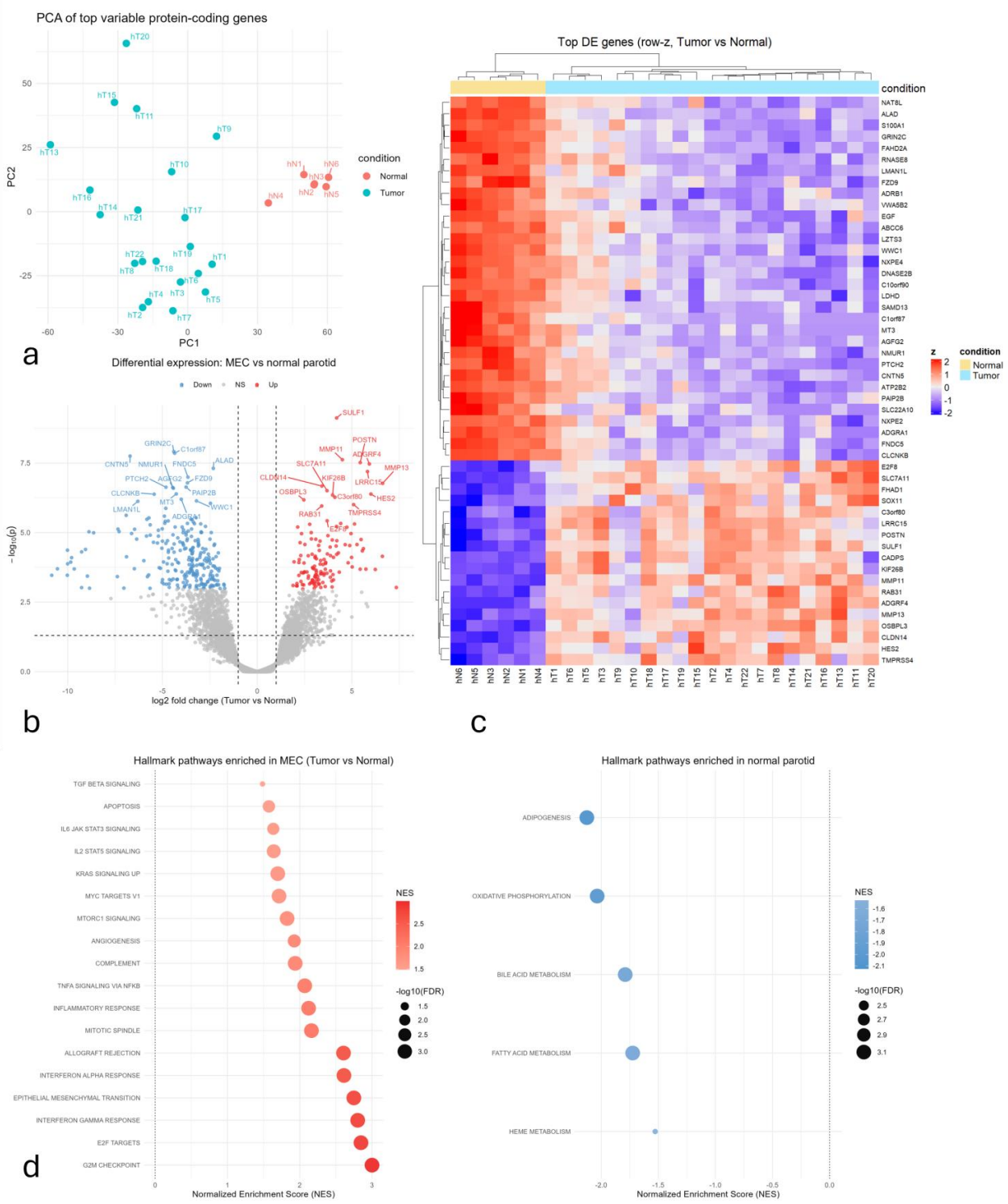

Figure 1: Transcriptomic and pathway differences between MEC and normal parotid tissue. **a:** PCA of the most variable protein-coding genes. **b:** Volcano plot of differentially expressed genes. **c:** Heatmap of representative differentially expressed genes. **d:** Hallmark enrichment analysis showing pathways enriched in MEC and normal parotid tissue. PCA, principal component analysis; MEC, mucoepidermoid carcinoma; FDR, false discovery rate; NES, normalized enrichment score

homeostatic pathways were relatively enriched in normal parotid tissue (Figure 1d).

The 27-gene union model yielded an AUC of 1.000 (95% CI, 1.000-1.000; Brier score, 0.036), whereas the compact three-gene model yielded an AUC of 0.976 (95% CI, 0.923-1.000; Brier score, 0.041). Given the small sample size and absence of an external validation cohort, these estimates should be interpreted as exploratory internal performance rather than evidence of generalizable diagnostic accuracy.

Predicted probabilities showed a clear separation between MEC and normal samples using the compact model (Figure 2a-b). POSTN had the largest positive penalized coefficient, MLXIPL had a negative coefficient, and SPP1 had a comparatively small coefficient in the final fitted model (Tables 1-2). The compact-panel genes also showed distinct expression patterns between MEC and normal parotid tissue (Figure 2c).

Discussion

Recent advances in tumor-microenvironment research have highlighted the interactions among malignant cells, stromal components, and immune cells [7, 16]. However, the molecular characteristics of salivary gland MEC remain insufficiently characterized, partly because of its rarity and biological heterogeneity. In the present study, four literature-derived MEC gene sets were integrated and evaluated for their ability to distinguish MEC from normal parotid tissue in GSE282430. The 27-gene union model and the reduced MLXIPL-POSTN-SPP1 candidate signature showed high internal discrimination, with AUCs of 1.000 and 0.976, respectively. Nevertheless,

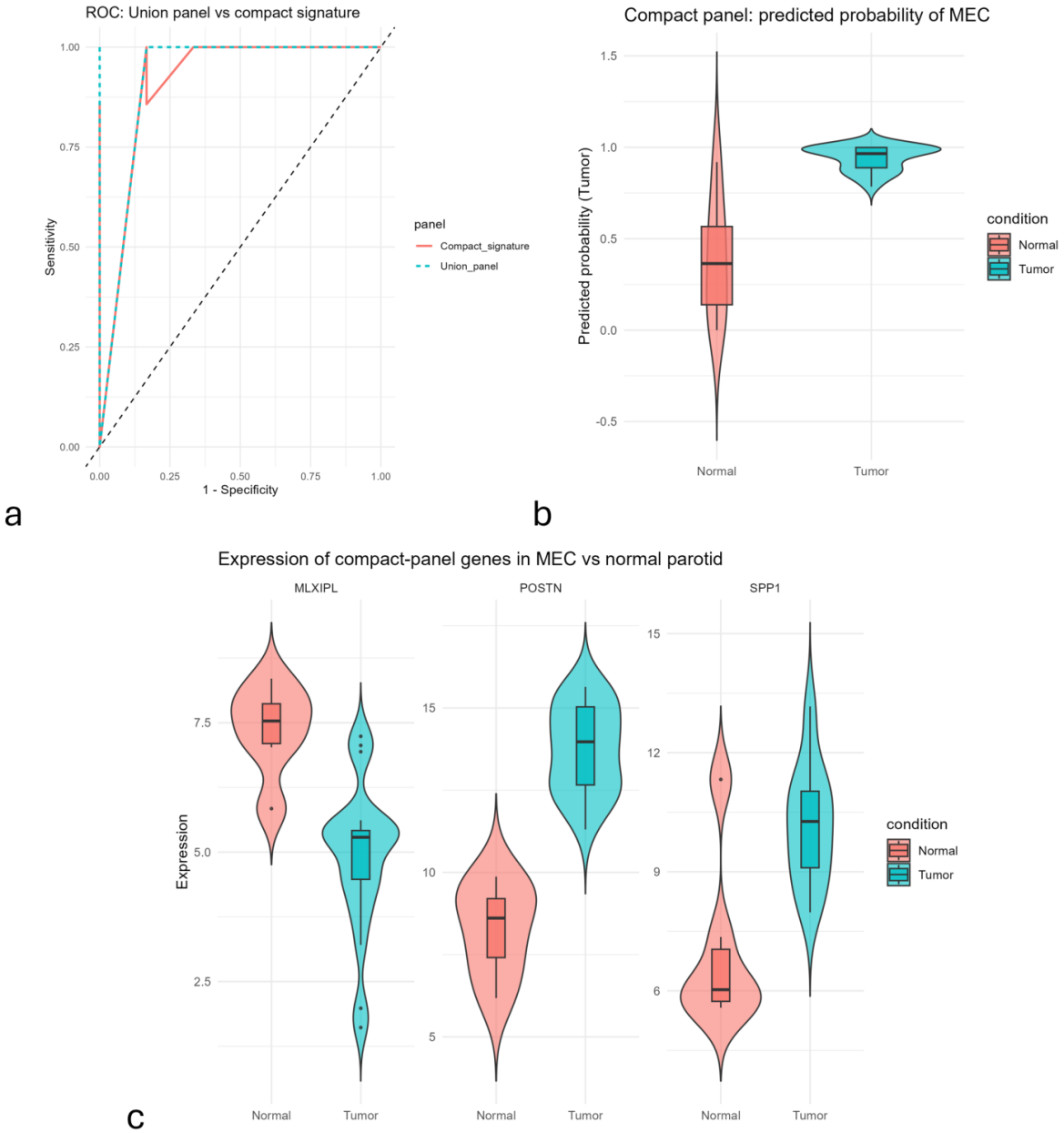

Figure 2: Tumor-versus-normal discrimination of the compact MLXIPL-POSTN-SPP1 exploratory model, **a:** Receiver operating characteristic curves comparing the 27-gene union panel and the compact three-gene signature, **b:** Predicted probabilities for MEC and normal samples using the compact model, **c:** Expression levels of MLXIPL, POSTN, and SPP1 across diagnostic groups. ROC: receiver operating characteristic; AUC: Area under the receiver operating characteristic curve; MEC: Mucoepidermoid carcinoma

Table 1: Tumor-versus-normal discrimination of literature-derived and compact MEC gene panels. AUC, 95% confidence intervals, and Brier scores are shown for each reconstructed gene panel, the 27-gene union panel, and the compact three-gene exploratory model

Panel	Number of genes	AUC	CI lower	CI upper	Brier
Naakka *et al.* [13]	11	0.960 [A]	0.881	1	0.104
Kang *et al.* [14]	9	0.566	0.267	0.866	0.186
Ramirez Martinez *et al.* [15]	2	0.801 [A]	0.527	1	0.131
Zou *et al.* [16]	6	1 [A]	1	1	0.030
Compact signature	3	0.976 [A]	0.923	1	0.041
Union panel	27	1 [A]	1	1	0.036

AUC: area under the receiver operating characteristic curve**; CI:** confidence interval**; MEC:** mucoepidermoid carcinoma;
[A]: Values sharing the same superscript letters are not statistically significant.

Table 2: Penalized coefficients of the final exploratory compact model

Term	Coefficient	Exponentiated coefficient
(Intercept)	-6.67854	0.001258
MLXIPL	-0.32267	0.724215
POSTN	0.854052	2.349146
SPP1	0.008549	1.008586

Penalized coefficients were obtained from the final LASSO model fitted to the complete cohort. Exponentiated coefficients are descriptive, depend on the scale of normalized gene expression, and should not be interpreted as conventional inferential odds ratios.

these estimates should be interpreted with caution because they were derived from a small, imbalanced cohort without external validation.

The concurrent selection of MLXIPL, POSTN, and SPP1 may reflect complementary metabolic, stromal, and immune-matrix alterations rather than a direct functional interaction among these genes. POSTN had the largest positive penalized coefficient in the final fitted model and was also included in the epithelial–mesenchymal transition-focused MEC gene set reported by Zou *et al.* [16]. POSTN encodes periostin, a secreted matricellular protein involved in extracellular-matrix organization and tissue remodeling. Studies in other tumor types have associated periostin with epithelial-mesenchymal transition, cellular invasion, stromal remodeling, and angiogenesis [17-18]. However, these findings cannot be directly extrapolated to MEC. The present bulk RNA-sequencing analysis also cannot determine whether POSTN expression originated predominantly from malignant epithelial cells, cancer-associated fibroblasts, or other stromal components.

SPP1 encodes osteopontin, an extracellular-matrix-associated phosphoprotein involved in cell adhesion and immune-stromal signaling. Osteopontin expression has previously been reported in salivary gland carcinomas, including MEC [19-20]. In the present analysis, SPP1 met the prespecified selection-frequency criterion but had a near-zero coefficient in the final full-cohort LASSO model. Its independent contribution at the selected regularization parameter was therefore limited, despite its recurrent selection across resampling folds. This distinction is important because selection frequency reflects the stability with which a gene enters resampled models, whereas the final penalized coefficient reflects its incremental contribution after accounting for the other retained genes. Accordingly, the inclusion of SPP1 should be interpreted as exploratory rather than as evidence that it is a dominant contributor to tumor-versus-normal discrimination.

MLXIPL showed a negative penalized coefficient, consistent with the reduced expression of lipid-metabolism-related genes reported by Kang *et al.* [14]. MLXIPL is a transcriptional regulator of carbohydrate-responsive and lipid-metabolic programs, including fatty-acid synthesis. Altered MLXIPL expression may therefore reflect metabolic reprogramming in MEC. Previous studies have linked lipid metabolism to immune regulation through effects on myeloid-derived suppressor cells, dendritic cell function, and immune surveillance [21-23]. However, the relationship between MLXIPL downregulation, lipid metabolism, and the immune microenvironment in MEC remains uncertain and requires functional investigation.

Collectively, the three-gene signature may capture a combination of reduced metabolic regulation and increased extracellular-matrix or stromal signaling. This interpretation remains hypothesis-generating because pathway activity, cellular origin, protein expression, and functional interactions were not directly assessed.

The comparator group consisted of histologically normal parotid tissue rather than benign or malignant salivary gland neoplasms. Consequently, the reported performance represents tumor-versus-normal discrimination and does not establish MEC-specific diagnostic accuracy. The signature may partly reflect broader neo-

plastic, stromal, inflammatory, or tissue-composition differences rather than alterations unique to MEC. Although histologic-grade information was available, the small numbers within individual grade groups did not permit reliable estimation of grade-specific performance. Therefore, the present study cannot determine whether the signature performs similarly across low-, intermediate-, and high-grade MEC.

Limitations

This study used a single public RNA-sequencing cohort containing 21 MEC and 6 normal parotid samples. The compact signature was derived and internally evaluated within this cohort; therefore, the observed performance may be optimistic and does not constitute external validation. Its reproducibility across institutions, populations, anatomical sites, expression platforms, and preprocessing pipelines remains unknown. The normal-tissue comparator also does not establish specificity relative to pleomorphic adenoma, Warthin tumor, adenoid cystic carcinoma, acinic cell carcinoma, or other salivary gland lesions. Although histologic grade information was available, the small grade-specific subgroups did not permit reliable evaluation of low-, intermediate-, and high-grade MEC. Finally, the findings are based on bulk transcriptomic data and require independent validation and experimental confirmation at the RNA and protein levels.

Conclusion

In conclusion, integration of four literature-derived gene sets identified an exploratory MLXIPL-POSTN-SPP1 candidate expression signature that distinguished MEC from normal parotid tissue within a small public RNA-sequencing cohort. Its specificity for MEC, performance across histologic grades, cross-platform reproducibility, and clinical relevance require independent validation in larger cohorts containing benign and malignant salivary gland comparator lesions.

Acknowledgments

The authors would like to thank all individuals and institutions who contributed, directly or indirectly, to this work. AI software was only used to correct grammatical errors and improve the fluency of the text to align it with native English standards.

Funding

No funding was received for this research.

Ethics Statement

Not Applicable

Conflict of Interest

None

References

[1] Yoshida EJ, García J, Eisele DW, Chen AM. Salivary gland malignancies in children. Int J Pediatr Otorhinolaryngol. 2014; 78: 174-178.

[2] De Felice F, de Vincentiis M, Valentini V, Musio D, Mezi S, Lo Mele L, et al. Management of salivary gland malignant tumor: The Policlinico Umberto I, "Sapienza" University of Rome Head and Neck Unit clinical recommendations. Crit Rev Oncol Hematol. 2017; 120: 93-97.

[3] Sood S, McGurk M, Vaz F. Management of Salivary Gland Tumours: United Kingdom National Multidisciplinary Guidelines. J Laryngol Otol. 2016; 130: S142-S149.

[4] De Felice F, Valentini V, De Vincentiis M, Di Gioia CRT, Musio D, Tummolo AA, et al. Prediction of Recurrence by Machine Learning in Salivary Gland Cancer Patients After Adjuvant (Chemo)Radiotherapy. In Vivo. 2021; 35: 3355-3360.

[5] de Oliveira FA, Duarte EC, Taveira CT, Máximo AA, de Aquino EC, Alencar Rde C, et al. Salivary gland tumor: a review of 599 cases in a Brazilian population. Head Neck Pathol. 2009; 3: 271-275.

[6] Israel Y, Rachmiel A, Ziv G, Nagler R. Benign and Malignant Salivary Gland Tumors - Clinical and Demo graphic Characteristics. Anticancer Res. 2016; 36: 4151-4154.

[7] Walsh H, Alghamdi S, Dave M, Alsanie I, Khurram SA. Diagnostic and prognostic biomarkers in salivary gland tumours. Diagnostic Histopathology. 2023; 29: 177-187.

[8] Gnjatic S, Bronte V, Brunet LR, Butler MO, Disis ML, Galon J, et al. Identifying baseline immune-related biomarkers to predict clinical outcome of immunotherapy. J Immunother Cancer. 2017; 5: 44.

[9] Mlecnik B, Bindea G, Angell HK, Maby P, Angelova M, Tougeron D, et al. Integrative Analyses of Colorectal Cancer Show Immunoscore Is a Stronger Predictor of Patient Survival Than Microsatellite Instability. Immunity. 2016; 44: 698-711.

[10] Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, et al. The Immune Landscape of Cancer. Immunity. 2019; 51: 411-412.

[11] Linxweiler M, Kuo F, Katabi N, Lee M, Nadeem Z, Dalin MG, et al. The Immune Microenvironment and Neoantigen Landscape of Aggressive Salivary Gland Carcinomas Differ by Subtype. Clin Cancer Res. 2020; 26: 2859-2870.

[12] Persson M, Andrén Y, Mark J, Horlings HM, Persson F, Stenman G. Recurrent fusion of MYB and NFIB transcription factor genes in carcinomas of the breast and head and neck. Proc Natl Acad Sci U S A. 2009; 106: 18740-18744.

[13] Naakka E, Barros-Filho MC, Adnan-Awad S, Al-Samadi A, Marchi FA, Kuasne H, et al. miR-22 and miR-205 Drive Tumor Aggressiveness of Mucoepidermoid Carcinomas of Salivary Glands. Front Oncol. 2021; 11: 786150.

[14] Kang H, Seo MK, Park B, Yoon SO, Koh YW, Kim D, et al. Characterizing intrinsic molecular features of the immune subtypes of salivary mucoepidermoid carcinoma. Transl Oncol. 2022; 24: 101496.

[15] Ramírez-Martínez CM, Jacinto-Alemán LF, Cruz-Hervert LP, Portilla-Robertson J, Leyva-Huerta ER. Bioinformatic Analysis for Mucoepidermoid and Adenoid Cystic Carcinoma of Therapeutic Targets. Vaccines (Basel). 2022; 10: 1557.

[16] Zou YP, Shan XF, Qiu JX, Geng Y, Xie S, Xiang RL, et al. Systematic identification of pathological mechanisms, prognostic biomarkers and therapeutic targets by integrating lncRNA expression variation in salivary gland mucoepidermoid carcinoma. Sci Rep. 2025; 15: 1573.

[17] Yan W, Shao R. Transduction of a mesenchyme-specific gene periostin into 293T cells induces cell invasive activity through epithelial-mesenchymal transformation. J Biol Chem. 2006; 281: 19700-19708.

[18] Siriwardena BS, Kudo Y, Ogawa I, Kitagawa M, Kitajima S, Hatano H, et al. Periostin is frequently overexpressed and enhances invasion and angiogenesis in oral cancer. Br J Cancer. 2006; 95: 1396-1403.

[19] Bjørndal K, Larsen SR, Godballe C, Krogdahl A. Osteopontin expression in salivary gland carcinomas. J Oral Pathol Med. 2011; 40: 451-455.

[20] Fok TC, Lapointe H, Tuck AB, Chambers AF, Jackson-Boeters L, Daley TD, et al. Expression and localization of osteopontin, homing cell adhesion molecule/CD44, and integrin αvβ3 in mucoepidermoid carcinoma and acinic cell adenocarcinoma of salivary gland origin. Oral Surg Oral Med Oral Pathol Oral Radiol. 2014; 118: 320-329.

[21] Yan D, Adeshakin AO, Xu M, Afolabi LO, Zhang G, Chen YH, et al. Lipid Metabolic Pathways Confer the Im munosuppressive Function of Myeloid-Derived Suppressor Cells in Tumor. Front Immunol. 2019; 10: 1399.

[22] Jiang L, Fang X, Wang H, Li D, Wang X. Ovarian Cancer-Intrinsic Fatty Acid Synthase Prevents Anti-tumor Immunity by Disrupting Tumor-Infiltrating Dendritic Cells. Front Immunol. 2018; 9: 2927.

[23] Tiwary S, Berzofsky JA, Terabe M. Altered Lipid Tumor Environment and Its Potential Effects on NKT Cell Function in Tumor Immunity. Front Immunol. 2019; 10: 2187.