ISSN: 3079-6601

Metabolomics-Based Discovery of Biomarkers for Oral Squamous Cell Carcinoma and Their Clinical Utility

Zhou Tianyi[1], Xu Jiayi[2], Zhou Wenjing[2], Wang Mengge[2], Li Ying[2]

[1]School of Basic Medical Sciences, University of South China, Hengyang 421001, China

[2]College of Stomatology, Hunan University of Medicine, Huaihua 418000, China

Abstract

The most common pathological type of the oral malignancies is oral squamous cell carcinoma (OSCC). Most patients visit the doctor when they notice signs like pain, ulceration, or dysphagia, and as such, patients have not been identified as soon as possible and recurrence and lymph node metastasis have formed the basis of survival. Metabolomics, which concentrates on small-molecule metabolites, has the potential to generate comprehensive >single cell-type phenotypic readout of tumor cell metabolic reorganization, host immune reactions, nutrition, and microecological disturbances on biofluids and tissues, thereby providing translatable leads in noninvasive screening, risk stratification, and longitudinal follow-up herein OSCC. As a strategy, this article consists of synthizing available evidence as well as systematically re-presenting resources of published findings, providing the discovery pathway of OSCC metabolic biomarkers based on sample types (plasma/serum/saliva/tissue), analytical platforms (NMR and mass spectrometry) and statistical modeling (multivariate discrimination and targeted validation). Quantitative findings directly presented in three published papers are compiled into re-usable graphical evidence: subgroup effects on ROC performance (AUC, sensitivity, specificity) of a plasma multivariate model based on NMR; monotonic fluctuations of key serum metabolites between stages I and IV (Lu et al., 2025); AUC ranking of various biomarkers of selected metabolites in tumor tissue (Zhou et al., 2020). Taken together, the evidence indicates that evidence of OSCC metabolic biomarkers can be extrapolated beyond " case control discrimination to staging aggressiveness stratification in spite of more stringent cohort matching coupled with confounding controls, cross-platform reproducibility, more rigorous targeted quantification workflows and multicenter prospective validation models.

Keywords

Oral squamous cell carcinoma, Metabolomics, Liquid biopsy, Staging, ROC, Lipid metabolism

1. Introduction

The history of clinical management of OSCC has been a historically rigid structural tension histopathology as the diagnostic gold standard that would require invasive tissue sampling, and prognosis through the early diagnosis and clear stratification, the information that is most demanded when the symptoms are nonclassical and early visual examination demonstrations are inconclusive. The local environment that is exposed to the mouth cavity closely associates OSCC with smoking, alcoholism, chewing betel quids, chronic inflammation, periodontal conditions, and oral microbiome. It is these aspects that affect the carcinogenesis and redefine the metabolomics in biofluids and tissues, i.e. that easily quantifiable indicators also important confounding signals. The metabolomics value is in the fact that it does not only represent a

ISSN: 3079-6601

single point change in a particular gene or protein but is a reflection of what happens on a systems level, regarding the overall impact of the tumor growth, immune response, oxidative stress, energy metabolism, and membrane lipids synthesis. This is a phenotype-proximal property that causes metabolomic readouts to be more similar to the interpretable outputs necessary to a clinical decision-maker. Saliva has a natural benefit in being a noninvasive oral cancer specimen, plasma/serum capture captures more systemic metabolic disturbances, and tissue metabolomics offers a framework to be interpreted mechanistically and a panel to be chosen. In case it is possible to make consistent pathway fingerprints between samples matrices, the robustness of biological markers and generalization are significantly enhanced (Lohavanichbutr et al., 2018; Nijakowski et al., 2022). In keeping with the best practices in journalism, this article focuses on the state-of-the-art evidence-chain organization: it constructs graphical evidence on the basis of published outcomes with clearly reported values and comments on clinical value and implementation requirements as well as the grounding logic of discovery-validation-translation (which is the major fallacy).

2. Literature Review

Studies on OSCC metabolomics have been able to advance in two major directions. One of the paths is centered around biofluids and focuses on the deployability of noninvasive or minimally invasive testing. LCM-MS apparatus is often used in salivary research to develop the differential metabolite profile and report changes in amino acids, organic acids, lipids, and molecules related to oxidative stress and subsequently assess diagnostic performance with training/validation schemes. Lohavanichbutr et al. (2018) showed that the discriminatory ability of metabolomic signatures between oral cancer and controls could be high in a comparatively large cohort of saliva, which highlights the necessity of addressing demographic factors and batch effects in a real-world population. Systematic reviews also show that salivary metabolomics has an overall potential in the diagnosis of OSCC, but there is still a high level of heterogeneity across studies that has been caused by variations in sample collection and preprocessing, platform coverage, statistical plans, control matching, and confounder adjustment (Nijakowski et al., 2022). The alternative pathway involves the emphasis on tissues and blood, and provides the systems-wide manifestation of tumor metabolic restructuring and the capability to target the validation. Tissue metabolomics has the capacity to better localize tissue pathway perturbations that are also of interest to tumor cell biology, and once there it can push candidate markers into blood or saliva to be further validated, so that a closed mechanism-to-detection loop can be formed. On plasma/serum metabolite studies, NMR would be the best choice to use in multicenter studies since its reproducibility and quantitative stability are high, and mass spectrometry would be the best choice in terms of coverage depth and pathway resolution. Plasma NMR metabolomics with metabolite set enrichment analysis was used to show diagnostic ability of a multivariate model and to return ROC results stratified by N stage and T stage; so-called clinical subgroup performance is more likely to be useful in real-life settings than a single overall AUC as patients are distributed differently across different screening, metastasis risk identification, and post-treatment monitoring settings. Lu et al. (2025) highlighted stage-specific metabolic characteristics of serum metabolomics, by demonstrating that some lipid molecules have monotonic manifestations with changes according to stage; these trend signals provide an entrance point with more clinical attraction to stratify risks and monitor over time than a dichotomous framing of cancer versus non-cancer. By integrating untargeted and targeted (Zhou et al., 2020), the authors could present tissue-level AUCs of various amino-acid biomarkers and hence offer a quantitatively based prioritization of the discovery to targeted validation. In general, the available literature repeatedly links the dysregulation of lipid metabolism (glycerophospholipids and sphingolipids) and amino acid metabolism. In order to produce high-quality journals and

ISSN: 3079-6601

clinical translation, the design of studies must ensure that they solve three fundamental problems: high control and like confounder controls, cross platforms and cross center reproducibility, and the interpretability and verifiability of the path between clinical model decision thresholds and statistical models.

3. Methods

In this article, the evidence-synthesis writing structure has been used. Instead of producing new primary experimental evidence, it restructures, so-called verifiable quantitative findings undertaken in published literature into reusable forms of clinical evidence and based on it, gives research-design recommendations with translationive orientations. Only those studies underwent inclusion when the study population was made known as OSCC; when they had numerical results, which could be directly cited (e.g., AUC, sensitivity, specificity, stage-specific mean); and when they were directly relevant to clinically-relevant activities (diagnostic discrimination, stage-specific mean).

Three types of figures are intended to be visualized and are represented as chart-generation prompts intended to be used with nano-banana and other similar tools. This is so such that all values recorded in each figure are all taken as required in the published source and no inferential filled-in figures are added. This figure, Figure 1, is derived on subgroup AUC/sensitivity/specificity, a plasma NMR multivariate ROC (Polachini et al., 2023). Using reported mean changes in serum metabolites at each of the stages I-IV, Figure 2 represents the results (Lu et al., 2025). Figure 3 depicts the ranking of amino-acid biomarkers of targeted tissue metabolomics of AUC (Zhou et al., 2020). There is no extrapolation of what is not reported in the original articles to those areas where statistical interpretation occurs strictly, with no extrapolation of thresholds, confidence intervals, and effect sizes. All Conclusions in the Discussion are also a distinction between what are evidence-supported observations and those that are reasonable mechanistic inferences, in line with the best practices at journal standards of evidence grading and disclosure of inference.

4. Results

The plasma metabolomic multivariate model shows that OSCC cannot be equally separated in all clinical subpopulations, which is especially important due to the early screening and patient stratification. AUC of 0.853 was reported by Polachini et al. (2023), when one is compared to discriminating overall OSCC cases against controls, the further results were stratified by nodal status and tumor T stage: 0.811 N0 vs control and 0.796 N+ vs control; 0.757 T1–T2 vs control and 0.872 T3-T4 vs control. Each comparison was annotated with sensitivity and specificity (Se/Sp: overall: 0.74/0.85; N0: 0.79/0.88; N +: 0.68/0.90; T1–T2: 0.64/0.84; T3 -T4: 0.88/0.93). A practical implication of this is that higher tumor burden subgroups are likely to have more pronounced systemic metabolic perturbations, with the model more able to capture separability, but the early-stage T1T2 sub-group where noninvasive tools may most be required have a performance more similar to the boundary of real-world applicability. Using the gross AUC as an indicator of appropriateness in early screening would hence overestimate the performance in clinical conditions that would be heavily case-mixed, and it is critical to record subgroup performance as a routine characteristic of OSCC metabolomic biomarker investigations and not as a supplementary measure.

ISSN: 3079-6601

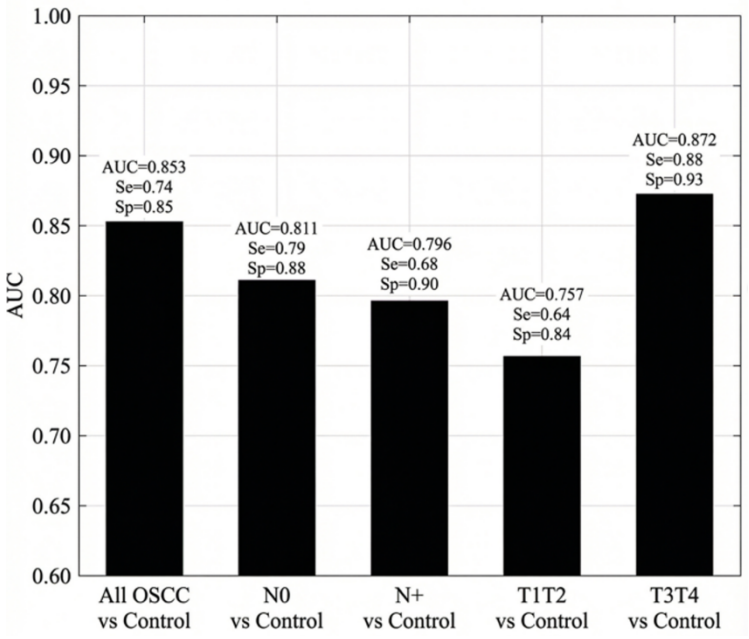

Figure 1. Multivariate ROC performance across OSCC clinical subgroups
Data: Polachini et al., 2023, Scientific Reports

Trend evidence, which is dependent on stage, changes the emphasis on diagnosis to one of stratification and monitoring. In one of such serum metabolomics experiments, Lu et al. (2025) showed mean values of several metabolites in OSCC stage I-IV and related stage-specific phenotypes via systemic changes like the metabolism of glycerophospholipids. Representative molecules had definite monotonic trends: the average LPC(20:4) met an increment of 0.005600137 to the 0.29020743 in stage I and stage IV respectively, and PC(40:6) met an increment of 0.085821925 to 0.29020743 in stage I and IV respectively; whereas of Cer(44:1) was decreasing towards 0.031746984 This type of stage-dependent monotonic signal is most consistent with clinical pathways since staging is a cumulative accumulation of change in tumor progression and systemic accumulation of metabolic burden; in cases where metabolite directional changes appear small, they can be used to support risk ranking or longitudinal follow-up surveillance instead of assigning the entire weight of the translatonal approach to a single-time-point dichotomous diagnostic decision.

ISSN: 3079-6601

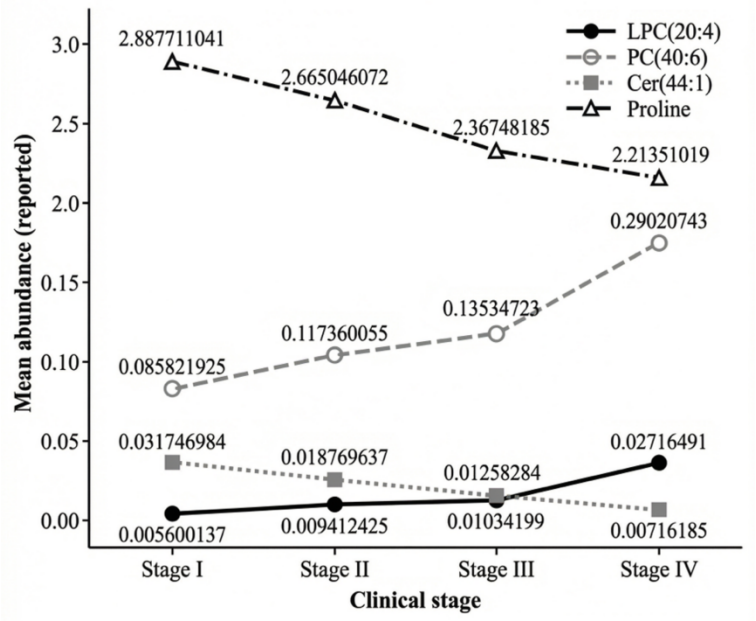

Figure 2. Monotonic serum metabolite trends across OSCC stages (I-IV)
Data: Lu et al., 2025, BMC Oral Health (Table values).

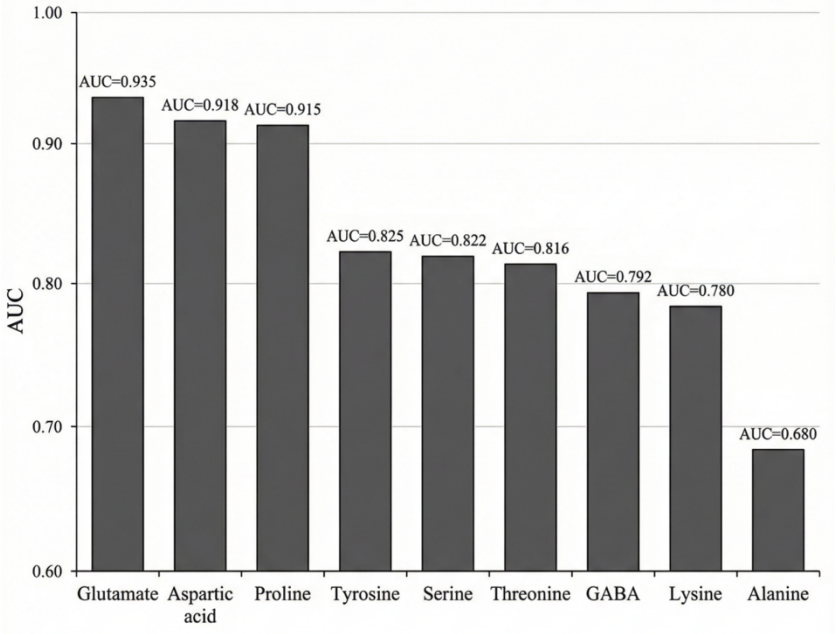

Figure 3. AUC ranking of amino-acid biomarkers
Data: Zhou et al., 2020, Frontiers in Oncology

The superior indicators of candidate panels verifiability Tissue-level targeted evidence are the more convincing quantitative indicators of verifiability. After targeted analysis on tissue samples by UHPLC master-slave methods using MS/MS techniques, Zhou et al. (2020) have reported the AUCs of various metabolites of the different amino-acids that can distinguish between tumor tissue and normal tissue, where glutamate (UC of 0.935), aspartic acid (UC of 0.918) and proline (UC of 0.915) became the highest. When it is desired to use AUC as a direct indicator to prioritize targeted validation, then it is possible to assign the metabolites as a high-priority candidate. The strength of tissue-based and biofluid-based biomarkers: The pathway level matching of tissue-based and biofluid-based biomarkers can be used to further enhance

ISSN: 3079-6601

panel robustness: When a pathway is dysregulated in tissue and has a concordant systemic perturbation in blood or saliva, panel candidates have an enhanced chance of replicating across centers and platforms. On the other hand, a biofluid metabolite which seems relevant in one small cohort and has no mechanistic explanation at the tissue level is more likely to be caused by a short-term change in diet, inflammation, or oral microecology, and thus is more of a high translational risk.

5. Discussion

Collectively, the evidence all comes down to a translationally critical nexus: OSCC metabolomic biomarkers exhibit statistical separability in a highly clinic-dependent and case-mix distribution way. Researchs that monitor an overall AUC without considering the relevant clinical subgroups are not expected to provide answers to the questions that are of interest in practice. This better performance of the plasma NMR model in subgroup of T3 4 is not surprising, as, higher tumor burden usually implies better-field inflammation and catabolic stress and consequent, greater systemic metabolic perturbations. The patients that need noninvasive support the most are individuals with the initial or territorial stages of the disease, and it can be assumed that the fact that the AUC is relatively lower in the T1-T2 subgroup means that future research must not resort to the standard paradigm on healthy controls. These controls should be specified in a manner that is more reflective of actual implementation, e e.g., oral potentially malignant disorders (OPMD) cohort or high-risk cancer-free people and where a confounder-control framework is prespecified to include smoking, alcohol consumption, periodontal inflammation, medication history, body mass index, time of sampling and state of diet. In the absence of these design decisions, positive predictive value may decrease significantly in real clinical practices due to variations in the prevalence of diseases and the confounding design and the leading journals will assume that the results are subject to a classic internally valid but externally uncertain risk of overfitting.

The advantage of monotonic stage-related trends is that they transform metabolomics into a no-once diagnostic instrument to a longitudinal disease-management instrument. The increase in lipid molecules with the stage can indicate increased demands of the membrane biogenesis and signaling lipids, and the reduction of ceramides can be associated with the tumor-driven remodeling of the sphingolipid homeostasis and decreases in proline can be linked with coveting amino-acid supplies, collagen-linked metabolism, and microenvironmental rearrangements. Such descriptions are biologically reasonable, but high-impact journals generally insist on the distinction between things seen and things theorized about and urges the confirmation of important nodes by multi-omics integration or functional experiments. In a more prosaic translationational approach, stage-linked molecules should be positioned as companion indicators of risk tiering in clinical practice: to announce aggressiveness before treatment, monitor rebounds on trends after treatment, and corroborative evidence and imaging or pathological reevaluation during follow up. The benefit of long-term use is that patients can act as their own controls, which decreases noise in inter-individual variability in diets and microecology, which has been particularly significant with the highly exposed oral environment in mind. Tissue-based targeted AUC ranking provides a candidate-selection strategy, which is closer to clinical lab medicine. The UHPLC-MS/MS is more susceptible to standardization and quality control at the hospital lab, and when combined with stable isotope-labeled internal standards, it can be audit-quantitative workflows. When research metabolomics is translated to regulated in vitro diagnostics, studies that have reduced its candidates to a small list of high-AUC molecules and provide clear analytical-performance measures (linearity, limits of detection/quantification, within- and between-run variability and stability) are possible. Simultaneously, metabolite annotation and naming consistency are

ISSN: 3079-6601

usually regarded as weaknesses of metabolomics; the highest level of journals generally requires compliance with high standards of annotation and structural verification based on authoritative databases (Wishart et al., 2022; Pang et al., 2021). The lack of transparency in annotations compromises the cross-study comparability and this compromise the reproducibility of panel development.

Clinical utility may be plotted on three different activities; triage screening in people of high risk, risk stratification of known patients, and relapse tracking in post-treatment. The tasks have varying demands on sensitivity, specificity, thresholds and cost functions, thus it is impractical to assume that a single panel can perform in the same way in all the uses. A more realistic design would be a tiered panel architecture where a small panel of stable metabolites would form a base panel of triage screening, and a larger panel of lipids and tissue-mechanism-supported molecules can be used as the stratification and monitoring. Clinical net benefit demonstrated at the modeling level should be presented by prespecified thresholds, external validation and analysis of decision-curves instead of the use of AUC alone, as the only determinant of value.

6. Conclusion

Metabolomics provides a pathogenically impactful directive to interlink OSCC biology and translationally significant choice as metabolic receivers considering that tumor-instigated reprogramming and host-based statuses of inflammatory and nutrition merge nutritionally and inflammatory. The results of evidence synthesis of published quantitative findings suggest that diagnostic performance is different across clinical subgroups, several serum metabolites show monotonic stage-related changes that could be used as stratification and longitudinal markers, and tissue-specific amino-acid markers are high-AUC with the potential to support prioritized validation. However, the main requirement of translational readiness is rigorous cohort matching and confounder control, standardized pre-analytics and targeted quantification, clear metabolite annotation, and multi-center prospective validation with pre-specified thresholds and utilities based on clinical grounds.

References

[1] Lohavanichbutr, P., Zhang, Y., Wang, P., Gu, H., Nagana Gowda, G. A., Djukovic, D., Buas, M. F., Raftery, D., & Chen, C. (2018). Salivary metabolite profiling distinguishes patients with oral cavity squamous cell carcinoma from normal controls. PLOS ONE, 13(9), e0204249. https://doi.org/10.1371/journal.pone.0204249

[2] Lu, W., Liu, L., Li, J., Yang, H., Yang, Y., Hu, Q., Song, X., Liu, L., & Yang, X. (2025). Serum metabolomics reveals distinctive signature of cancer staging in oral squamous cell carcinoma. BMC Oral Health, 25, Article 1826. https://doi.org/10.1186/s12903-025-07210-9

[3] Nijakowski, K., Gruszczyński, D., Kopała, D., & Surdacka, A. (2022). Salivary metabolomics for oral squamous cell carcinoma diagnosis: A systematic review. Metabolites, 12(4), 294. https://doi.org/10.3390/metabo12040294

[4] Pang, Z., Chong, J., Zhou, G., Morais, D. A. D. L., Chang, L., Barrette, M., Gauthier, C., Jacques, P.-É., Li, S., & Xia, J. (2021). MetaboAnalyst 5.0: Narrowing the gap between raw spectra and functional insights. Nucleic Acids Research, 49(W1), W388–W396. https://doi.org/10.1093/nar/gkab382

[5] Polachini, G. M., Meirelles, L. M., Castro, L. E. V., Bellini, M. H., & Kowalski, L. P. (2023). Plasma metabolomics of oral squamous cell carcinomas by NMR spectroscopy and metabolite set enrichment analysis. Scientific Reports, 13, 8588. https://doi.org/10.1038/s41598-023-34808-2

[6] Wishart, D. S., Guo, A., Oler, E., Wang, F., Anjum, A., Peters, H., … & Gautam, V. (2022). HMDB 5.0: The Human Metabolome Database for 2022. Nucleic Acids Research, 50(D1), D622–D631. https://doi.org/10.1093/nar/gkab1062

ISSN: 3079-6601

[7] Zhou, Y., Wang, M., Li, H., Shi, L., & Zhang, T. (2020). Integrated non-targeted and targeted metabolomics uncovers amino acid markers of oral squamous cell carcinoma. Frontiers in Oncology, 10, 426. https://doi.org/10.3389/fonc.2020.00426