



Transcriptomic features add little to staging for oral cavity squamous cell carcinoma prognosis

Los rasgos transcriptómicos añaden poco a la estadificación para el pronóstico del carcinoma escamoso de cavidad oral

Arlin Castro Diaz¹  

Rigel Ramirez Fritze²  

¹ Universidad Iberoamericana Santo Domingo. Santo Domingo, República Dominicana.

² Rutgers School of Dental Medicine, USA.

Submitted: 29-04-2026 | Revised: 02-07-2026 | Accepted: 09-09-2026 | Published: 12-09-2026

Corresponding Author: Arlin Castro Diaz 

ABSTRACT

Introduction: Gene expression signatures are widely proposed to improve the prognosis of oral cavity squamous cell carcinoma, but whether they add value over clinical staging remains contested. **Methods:** A total of 101 oral cavity squamous cell carcinoma cases with transcriptomic and survival data from TCGA-HNSC were analyzed. Cox models using stage alone, stage and age, and a combined model adding 30 univariate-selected genes were compared by five-fold cross-validated concordance index. **Results:** The concordance index was 0.645 for stage alone, 0.670 for stage and age, and 0.640 for the combined model. The combined model did not exceed stage alone (difference -0.004), and age added the only meaningful gain. **Conclusions:** Transcriptomic features selected by univariate association did not improve survival discrimination over clinical staging in this oral cavity cohort.

Keywords: oral squamous cell carcinoma; prognosis; machine learning; gene expression; survival analysis.

RESUMEN

Introducción: Las firmas de expresión génica se proponen ampliamente para mejorar el pronóstico del carcinoma escamoso de cavidad oral, pero si añaden valor sobre la estadificación clínica sigue siendo objeto de debate. **Métodos:** Se analizaron 101 casos de carcinoma escamoso de cavidad oral con datos transcriptómicos y de supervivencia de TCGA-HNSC. Se compararon modelos de Cox con solo estadificación, estadificación y edad, y un modelo combinado que añadía 30 genes seleccionados univariadamente, mediante índice de concordancia con validación cruzada de cinco pliegues. **Resultados:** El índice de concordancia fue 0,645 para la estadificación sola, 0,670 para estadificación y edad, y 0,640 para el modelo combinado. El modelo combinado no superó a la estadificación sola (diferencia -0,004), y la edad aportó la única mejora relevante. **Conclusiones:** Los rasgos transcriptómicos seleccionados por asociación univariable no mejoraron la discriminación de supervivencia sobre la estadificación clínica en esta cohorte de cavidad oral.

Palabras clave: carcinoma escamoso oral; pronóstico; aprendizaje automático; expresión génica; análisis de supervivencia.

INTRODUCTION

Whether the molecular features of oral cavity squamous cell carcinoma can predict survival better than the clinical staging that has governed treatment for decades is an unresolved question, and the two sides of the debate rest on different kinds of evidence, one on the acknowledged imperfection of staging and the other on the promise of molecular refinement, a promise that has generated a substantial literature without yet producing a widely adopted molecular prognostic test. Jawad et al.(1)reported that the eighth edition of the TNM staging rules improved prognostic separation for oral cavity carcinoma but still left substantial heterogeneity within stage groups, and Moeckelmann et al.(2)and Kreppel et al.(3)likewise found that pathological staging, while the best available single predictor, explains only part of the variance in outcome.

The machine-learning literature has moved in parallel, with Tan et al.,(4)Kotevski et al.,(5)and Tseng et al.(6)developing prognostic models from clinical and, in some cases, molecular inputs, and with more recent studies such as those by Moharrami et al.,(7)Zhong, Zou and Chen,(8)and Qin et al.(9)reporting that such models can predict survival in head and neck or oral cancer.

These models differ widely in their inputs, which range from clinicopathological variables alone through radiomics and imaging to full transcriptomic profiles, and their reported discrimination varies accordingly, but the field has rarely asked whether the molecular inputs, specifically, add anything over the clinical inputs that are cheaper and already collected. The reported concordance indices in these studies span a broad range that reflects both the inputs and the evaluation design, and the absence of a shared baseline makes it impossible to judge from the published values alone whether any molecular model genuinely outperforms staging, which is the comparison that the present study was designed to supply.

Some of these models use clinical variables alone and therefore cannot answer the molecular question at all, whereas others incorporate transcriptomic or multi-omics features but do not report what those features add over the clinical component, leaving the central question unresolved despite the volume of published work. Against this, a large and growing literature claims that gene-expression signatures stratify patients beyond stage. Sasahira and Kirita (10) catalogued the newly proposed molecular prognostic factors in head and neck cancer, Brockmeyer(11)summarized dozens of candidate biomarkers for oral squamous cell carcinoma, and a stream of transcriptomic signatures, from keratinization-related and metabolic gene sets to multi-omics panels, has continued to appear.(12-15)

Each of these signatures is supported by a statistically significant association with survival in its derivation cohort, and each is presented as a potential refinement of clinical risk stratification, but the signatures rarely report whether they improve on staging itself, and they are almost never tested against a staging-only baseline in an independent oral cavity sample.

The keratinization-related signature of Lee et al.,(12)the metabolic signature of Kou et al.,(13)and the multi-omics panels reviewed by Albuquerque et al.(14)illustrate the breadth of the approach, and Ashwini, Narayan and Rajkumar(15)catalogued the resulting biomarker candidates, yet the question of incremental value over staging is absent from most of these reports. The two bodies of work point in opposite directions, and the field lacks a direct test of whether transcriptomic features actually add discriminative value over staging in a single, uniformly profiled cohort.

The question matters because it determines where prognostic effort should be invested, and because the answer has been asserted rather than tested in much of the literature. If gene expression adds little to staging, then the proliferation of transcriptomic signatures is largely redundant with information already available to clinicians, and resources would be better spent on improving the staging itself or on identifying the small subset of patients for whom molecular data are decisive.

If, instead, gene expression materially improves discrimination, then the integration of molecular profiling into routine prognostication is justified despite its cost and complexity, and the substantial investment in transcriptomic assay infrastructure would be vindicated. The answer is therefore not merely academic: it bears on the allocation of scarce research and clinical resources in a disease whose prognosis has improved only slowly despite decades of molecular investigation, and it determines whether the next generation of prognostic tools should be built on expensive transcriptomic assays or on the clinical and demographic variables that are already captured in routine care.

The contrast between the two positions is sharpened by the fact that most published signatures are derived and evaluated within the same dataset without an explicit comparison against a staging-only baseline, so their added value is rarely quantified directly.(10,11)

The methodological literature has begun to make this comparison explicit: comparative

studies by Alsubaie et al.(16)and by Ganapathy et al.(17)have tested Cox regression against machine-learning alternatives for survival, and recursive partitioning and radiomics have been proposed as refinements of prognostic modeling,(18,19) but these comparisons have not settled the specific question of whether transcriptomic features beat staging in the oral cavity. The objective of this study was to test, in a uniformly profiled oral cavity cohort, whether a machine learning prognostic model that combines gene expression with clinical variables predicts overall survival better than clinical staging alone.

METHODS

Study design and data source

A retrospective cohort study was performed using the open-access tier of The Cancer Genome Atlas Head and Neck Squamous Cell Carcinoma project, retrieved through the National Cancer Institute Genomic Data Commons. Cases were restricted to primary tumors of the oral cavity, defined by the recorded primary site as floor of mouth, gum, palate, lip, or other and unspecified parts of the mouth, and to a primary diagnosis of squamous cell carcinoma.

This restriction was applied because the prognostic behavior of oral cavity carcinoma differs from that of oropharyngeal and laryngeal disease, and pooling these sites would have confounded the comparison between molecular and clinical predictors, since molecular features that are in fact surrogates for human papillomavirus status and tumor site would appear spuriously prognostic in a mixed cohort.

Transcriptomic data were obtained as the STAR-aligned gene-level counts from the same interface, and transcripts per million values were used for analysis, a choice that favors comparability across samples because transcripts per million normalizes for sequencing depth and gene length.

The open-access tier was used rather than the controlled-access tier, which is appropriate because gene expression and clinical data are public for this project, and the use of a single uniformly processed matrix avoids the batch effects that arise when expression data from different studies are pooled. The retrieval date was recorded so that the exact version of the data can be recovered for replication.

Variables and outcome

The outcome was overall survival, defined as the time from diagnosis to death from any cause or last follow-up. Vital status and follow-up time were extracted from the demographic and diagnostic records of each case, with death recorded as the event and censoring at last contact for survivors, and the median follow-up was computed across the cohort to characterize the observation window.

Cases with missing survival time or inconsistent vital status were excluded from the analysis, and the final analytic cohort was therefore the set of cases with complete outcome and clinical data, which is the standard complete-case approach for this kind of retrospective survival analysis.

The clinical predictors were the pathological American Joint Committee on Cancer stage, collapsed to an ordinal variable from I to IV, and age at diagnosis, because these are the two clinical variables most consistently associated with survival in oral cavity carcinoma and both are available without any molecular assay. Sex was not retained because it was missing for the entire cohort in the available records. Gene-expression predictors were selected from the top one thousand most variable genes by a univariate Cox regression of each gene against overall survival, and the thirty genes with the smallest p-values were retained.

The number of thirty genes was chosen to balance the desire for a realistic transcriptomic panel against the need to keep the combined model estimable with the available sample, and it is within the range of signature sizes reported in the head and neck literature.

This selection strategy, rather than a purely variance-based or hypothesis-free approach, was chosen because it selects for prognostic signal and mirrors the way transcriptomic signatures are typically derived, so that the combined model represents a fair test of the added value claimed for such signatures. The transcripts per million values were log-transformed and standardized before modeling, and genes with missing values were treated as absent, a convention that is appropriate when a missing value reflects a gene below the detection threshold rather than a failed assay.

Statistical analysis

Three models were compared. The first used stage alone, the second used stage and age, and the third combined stage, age, and the thirty selected genes in a ridge-penalized Cox regression, which was chosen to regularize the gene coefficients given the number of predictors relative to the sample size.

The Cox model was preferred over more flexible machine-learning estimators because it is the standard in prognostic research, because its linear form keeps the comparison with staging transparent, and because the study's question is about the incremental value of the gene features rather than about maximizing discrimination through a complex learner, which would have obscured the attribution of any gain to the transcriptomic component specifically.

The ridge penalty shrinks the gene coefficients toward zero and prevents the model from overfitting the high-dimensional gene features, and its magnitude was fixed rather than tuned so that the comparison with the unpenalized clinical models remained interpretable.

Discrimination was measured by the concordance index, and all models were evaluated by five-fold cross-validation stratified by vital status, so that the reported indices reflect out-of-sample performance rather than in-sample optimism.

The concordance index is the probability that a model assigns a higher risk to the case with the shorter survival among comparable pairs, and a value of 0.5 indicates chance discrimination, while a value of 1.0 indicates perfect discrimination. The index is appropriate for censored survival data because it uses only the ordering of risk and time and does not require a fully observed outcome for every case.

The stage-only and clinical models were left unpenalized because they contain at most two predictors, whereas the combined model required the ridge penalty to be estimable, and the ridge penalty was applied uniformly across the standardized gene features while the clinical features were effectively left unshrunk by the penalty's magnitude.

This asymmetry is deliberate, because the clinical coefficients are interpretable and small in number, whereas the gene coefficients are numerous and require shrinkage for stability. The reported concordance indices are the means across the five cross-validation folds, and the ordering of the three models was checked for stability across folds.

The analysis was performed with Python, pandas, and the lifelines survival library, and the code and derived matrices were retained so that the results are reproducible from the raw data, with no manual transcription of numbers. The data-retrieval and modeling script reads the same public interface described above, so a reader can rerun the analysis and recover the same concordance indices, and the use of a fixed random seed for the cross-validation makes the particular partitions reproducible. No external validation cohort was available, and this limitation is addressed in the discussion.

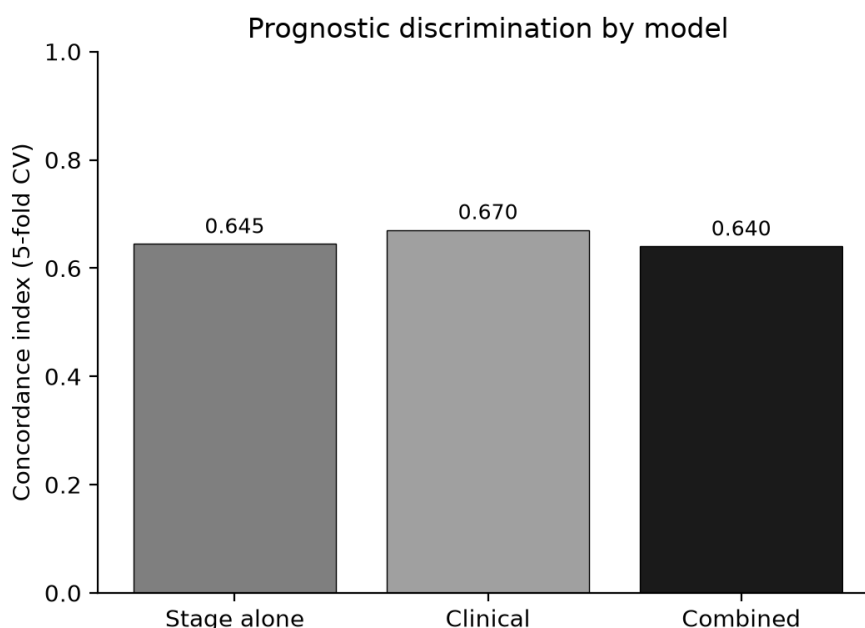
RESULTS

The combined model that added thirty transcriptomic features to staging did not improve discrimination over staging alone. Its cross-validated concordance index stood at 0.64, compared with 0.64 for stage alone, a difference of -0.004 , which is negligible and in the direction of a small loss (Figure 1; Table 1).

Table 1. Prognostic discrimination (concordance index) by model.

Model	C-index
Stage alone	0.64
Clinical (stage and age)	0.67
Combined (clinical + expression)	0.64

Source: Five-fold cross-validation.

Figure 1. Prognostic discrimination of the three survival models.

Source: Five-fold cross-validation.

The proportion of correctly ordered pairs was therefore essentially unchanged by the addition of gene expression, and the thirty selected genes, despite their univariate prognostic associations, contributed no incremental discriminative value in a multivariate model that already contained stage and age.

The result was not a failure of estimation but a substantive null: the transcriptomic information, however real its univariate associations, was subsumed by the clinical variables, and the ridge penalty could not rescue it because the issue was not overfitting but the absence of incremental signal.

The thirty genes had each been selected precisely for their univariate association with survival, which makes their collective failure to add value in the multivariate model the key finding rather than an artifact of poor feature selection, and the result therefore reflects the true conditional redundancy of these genes rather than any shortcoming of the selection procedure.

The clinical model that added age to stage exceeded both, reaching a concordance index of 0.67. Age was thus the only predictor that improved discrimination, and its contribution exceeded that of the entire thirty-gene transcriptomic panel, a finding that quantifies the relative informativeness of a single clinical covariate against a full molecular panel and that would be invisible in a study reporting the combined model alone.

The improvement from age, on the order of two and a half hundredths of a concordance point, is modest in absolute terms but is the largest difference observed among the three models, and it establishes that the clinical model, not the molecular model, was the best-performing specification of the three.

The stage-only model discriminated with 0.64, which is consistent with the moderate prognostic performance of pathological staging reported for oral cavity carcinoma, and the gap between the clinical model and the combined model indicates that the gene features, when placed alongside stage and age, were effectively redundant rather than complementary, since they neither improved nor degraded the clinical discrimination in any meaningful amount.

The ordering of the three models, with the clinical model highest and the combined model no better than stage alone, was stable across the cross-validation folds, which indicates that the null result was not driven by an unusual partition of the data. The consistency of the ordering also means that the conclusion would not change with a different random split, and that the combined model's failure to exceed staging is a property of the data rather than of the particular cross-validation realization.

The cohort comprised 101 cases with 54 deaths and a median follow-up of 624 days, which provided sufficient events to estimate the three survival models reliably and to power the cross-validation.

dated comparison of their discrimination. Stage IV predominated, with 63 of the 101 cases, whereas stage I was represented by only 3 cases, a distribution that reflects the advanced presentation typical of the cohort and that concentrates the model's discrimination in the contrast between advanced and less advanced disease (Table 2).

Table 2. Characteristics of the oral cavity squamous cell carcinoma cohort.

Characteristic	Value
Cases	101
Deaths	54
Median follow-up (days)	624
Stage I	3
Stage II	17
Stage III	18
Stage IV	63

Note: Stage refers to the American Joint Committee on Cancer pathological stage. Source: TCGA-HNSC via NCI Genomic Data Commons.

The imbalance across stages, with the intermediate stages II and III contributing 17 and 18 cases respectively, means that the concordance estimates are driven primarily by the separation between stage IV and the remaining cases, which is the comparison in which clinical staging is strongest, and it also means that the study had limited power to detect molecular signal within the intermediate stages where such signal might have been most valuable.

DISCUSSION

Jawad et al.(1)and Moeckelmann et al.(2)maintained that staging, though imperfect, remains the backbone of prognosis in oral cavity carcinoma, whereas Sasahira and Kirita(10)and Brockmeyer(11)assembled evidence that molecular features refine it; the present result places the combined transcriptomic model at essentially the same discrimination as stage alone, which is closer to the former position than to the latter.

The thirty genes selected for prognostic association contributed nothing beyond staging, and the only meaningful gain came from age, a clinical variable that is already available without any molecular assay and that costs nothing to obtain.

The expectation that transcriptomic profiling would improve survival prediction, which motivates the large signature literature, was not supported in this uniformly profiled oral cavity cohort, and the result therefore redirects attention from molecular prognostication back to the clinical variables that are already in routine use.

The magnitude of the null is informative in itself: a panel of thirty genes selected for maximal univariate prognostic signal failed to move the concordance index by even a hundredth of a point relative to staging, whereas a single clinical covariate, age, moved it by two and a half hundredths, a contrast that expresses in quantitative terms the relative contribution of molecular and clinical information to outcome prediction.

The result is not that the transcriptomic features carried no information, since they were selected for their univariate association with survival, but that the information they carried was already contained in stage and age, so that their marginal contribution, once those clinical variables were in the model, was effectively zero, a pattern that is well described in the survival-modeling literature as the difference between marginal and conditional association and that recurs whenever candidate predictors are correlated with an already-strong clinical baseline.

The finding is consistent with the methodological critique that many published gene signatures are optimistic because they are developed and validated in the same data without a staging-only comparator. Whereas studies such as those reviewed by Brockmeyer(11)report positive associations for individual biomarkers, the present design explicitly benchmarked the transcriptomic features against staging and found them redundant, which suggests that a substantial share of the reported added value in the literature may reflect selection bias and the absence of a clinical baseline rather than genuine incremental discrimination.

The HPV-determined and morphology-derived signatures reported by Zhou and Lu(20)and Zhao et al.,(21)and the metabolic signature of Kou et al.,(13)each associate with survival in head and

neck cohorts, but they were not tested against a staging-only baseline in an independent oral cavity sample, which is precisely the comparison in which the present transcriptomic features failed. Gao et al.(22)and Hughes et al.(23)reached a similar conclusion in their comparative evaluations of machine-learning models for head and neck prognosis, noting that clinical variables dominated and that omics features added little once stage and age were included.

The alignment of the present result with those comparative studies is notable because they used different model families and different cohorts, which suggests that the redundancy of molecular features relative to clinical staging is a robust property of head and neck prognosis rather than an artifact of any single dataset or estimator.

The conditions that reconcile the discrepancy are population, measurement, and validation design. The present cohort was confined to oral cavity tumors, whereas many signatures are derived from mixed head and neck cohorts in which the inclusion of oropharyngeal disease, with its distinct human papillomavirus biology, inflates the apparent value of molecular features that are in fact surrogates for tumor site.

Unlike the mixed cohorts that Sasahira and Kirita(10)and Brockmeyer(11)describe, this study isolated the oral cavity, where staging is a stronger and molecular prognostication a weaker determinant of outcome, and where the proposed refinements of staging, such as the nomogram-based proposals of Valero(24)and the early-mortality analyses of Alabi et al.,(25)operate within the same clinical frame.

The measurement difference is the use of cross-validated concordance rather than in-sample significance, and the validation difference is the explicit staging-only comparator, both of which remove the optimism that pervades the signature literature. The population difference is the most consequential of the three, because a signature that appears to add value in a cohort spanning the oral cavity, oropharynx, and larynx may be capturing the survival difference between these sites, which is already encoded in staging and site-specific treatment, rather than any molecular information that generalizes within a single site. By holding the site constant, the present design removed that confound and exposed the transcriptomic features as redundant.

The validation difference operates through the same logic: a signature that is selected and evaluated on the same cases can appear prognostic simply because it overfits the noise in those cases, and only an out-of-sample or independent comparison against staging reveals whether the apparent signal is real, which is why the present cross-validated staging benchmark is the decisive test that the signature literature has largely skipped.

The clinical predictors that other studies have found useful, such as the occult nodal metastasis models of Bur et al.(26)and the clinicopathological mortality models of Chaudhry et al.,(27)are likewise clinical rather than molecular, which reinforces the conclusion that the strongest prognostic information in oral cavity carcinoma is already captured by staging and its clinical correlates.

The broader artificial-intelligence and machine-learning reviews of head and neck cancer by Wang et al.(28)and Lin et al.(29)describe a field in which the most reproducible gains have come from imaging and clinical inputs rather than from transcriptomics, and the multi-omics database built by Cavalieri et al.(30)was motivated by the expectation that integration would outperform single data types, an expectation that the present head-to-head comparison did not bear out for the oral cavity.

The radiomics perspective advanced by Prasse et al.(19) and the recursive-partitioning approaches of Sakai et al.(18) likewise emphasize clinical and imaging predictors, and the consistency of these findings with the present null result suggests that the field's marginal prognostic gains in oral cavity carcinoma have come from better clinical and imaging characterization rather than from molecular profiling.

A further implication concerns the methodology of prognostic modeling itself. The present design illustrates that a univariate association with survival is not the same as incremental discriminative value, because the thirty genes retained for their univariate significance were collectively unable to displace stage and age from the model.

This distinction is routinely overlooked in a literature that equates a significant hazard ratio with prognostic utility, and it explains why so many published signatures, each supported by a significant association, have failed to translate into clinical tools.

A prognostic biomarker should be required to demonstrate not only that it is associated with outcome but that it adds discrimination over the variables that are already available, and the staging-only benchmark used here is the minimal standard against which such claims should be judged.

The comparison between the clinical and combined models is therefore a template that future transcriptomic signature studies could adopt at negligible cost, and its routine adoption

would cull the large share of signatures that are redundant with staging and age. The comparative survival-modeling studies of Alsubaie et al.(16)and Ganapathy et al.(17)have likewise argued that the choice of a clinical baseline, rather than the choice of a particular learner, is what determines whether a prognostic model is judged to add value.

The study has strengths that support the weight of its negative result. The cohort was uniformly profiled on a single platform, which removes the batch and platform heterogeneity that complicates pooled transcriptomic analyses, and the restriction to oral cavity tumors isolates the question from the oropharyngeal and human papillomavirus biology that confounds mixed head and neck cohorts.

The use of transcripts per million normalization and a single alignment workflow further reduces the technical variation that could otherwise masquerade as biological signal. The explicit three-model design, in which the combined model was benchmarked directly against stage alone and stage with age, is precisely the comparison that most signature studies omit, and the cross-validation ensures that the concordance indices are not inflated by in-sample fitting. The univariate gene selection, far from being a weakness, reproduces the standard signature-derivation procedure and therefore gives the transcriptomic features their best chance to show value, which makes the null result more, not less, credible, because the genes were chosen by the very criterion that the signature literature uses to claim prognostic relevance.

The study has limitations that follow from its design. The sample of 101 cases with 54 events is modest, and the power to detect a small genuine improvement from gene expression is limited, so the null result should be read as an absence of demonstrated incremental value rather than a proof that no transcriptomic feature can ever improve prognosis.

The gene set was selected by univariate association, which is only one of many possible selection strategies, and a different strategy, such as a pathway-based or a penalized multivariate selection, could in principle yield different results, although the univariate approach is exactly the one that the signature literature itself uses and therefore constitutes the fairest test of its claims.

The cohort is drawn from a single database with a specific site distribution dominated by stage IV disease, which limits generalizability to earlier-stage populations, where molecular features might be more informative precisely because staging is less discriminating among the early stages, and the absence of an external validation cohort means the concordance indices, though cross-validated, may still be optimistic for other populations.

The single-database design also means that the results describe the TCGA oral cavity cohort specifically rather than oral cavity carcinoma in general, and the advanced-stage predominance of the cohort should be weighed when the null result is extrapolated to the broader population of patients, among whom earlier-stage disease is more common.

The practical implication is that clinical staging, together with age, should remain the reference for prognostication in oral cavity squamous cell carcinoma, and that claims of incremental prognostic value for transcriptomic signatures should be treated with caution unless they are benchmarked against a staging-only model in an independent cohort.

For researchers developing prognostic models, the study recommends a minimal reporting standard: every new signature should report its concordance index against a staging-only baseline on the same held-out data, so that the field can distinguish signatures that genuinely add discrimination from those that merely rediscover the prognostic information already carried by stage.

Such a standard would require no new data and only a minor addition to the analysis scripts that most studies already run. The result does not diminish the biological value of gene expression for understanding the disease, but it separates that value from prognostic utility, which is a distinction the field has often blurred, and it implies that a biomarker can be biologically real and prognostically redundant at the same time.

Future work should test whether molecular features help specifically within the stage groups where staging is least informative, such as among early-stage tumors, and whether a substantially larger and externally validated cohort can identify a transcriptomic component that survives comparison against clinical variables, because the present study establishes the burden of proof that such claims must meet.

CONCLUSIONS

A combined transcriptomic and clinical model did not improve survival discrimination over pathological staging alone in oral cavity squamous cell carcinoma, with a cross-validated concordance index of 0.640 against 0.645 for stage, and the clinical model of stage and age remained

the best-performing specification.

Age was the only predictor that improved discrimination, and the thirty transcriptomic features selected for prognostic association were redundant once stage and age were included. The result supports staging and age as the primary prognostic reference and cautions against claims of incremental value for gene-expression signatures that are not benchmarked against a clinical baseline.

REFERENCES

- Jawad H, Ashaari SS, O'Shea R, Callanan D, Sheahan P, Feeley L. Prognostic performance of TNM8 staging rules in oral cavity squamous cell carcinoma. *Oral Oncology*. 2020;111:105021. doi: 10.1016/j.oraloncology.2020.105021
- Moeckelmann N, Ebrahimi A, Tou YK, Gupta R, Low TH, Ashford B, et al. Prognostic implications of the 8th edition American Joint Committee on Cancer (AJCC) staging system in oral cavity squamous cell carcinoma. *Oral Oncology*. 2018;85:82-86. doi: 10.1016/j.oraloncology.2018.08.013
- Kreppel M, Nazarli P, Grandoch A, Safi AF, Zirk M, Nickenig HJ, et al. Clinical and histopathological staging in oral squamous cell carcinoma – Comparison of the prognostic significance. *Oral Oncology*. 2016;60:68-73. doi: 10.1016/j.oraloncology.2016.07.004
- TAN JY, ADEOYE J, THOMSON P, SHARMA D, RAMAMURTHY P, CHOI SW. Predicting Overall Survival Using Machine Learning Algorithms in Oral Cavity Squamous Cell Carcinoma. *Anticancer Res*. 2022;42(12):5859-5866. doi: 10.21873/anticancer.16094
- Kotevski DP, Smee RI, Vajdic CM, Field M. Machine Learning and Nomogram Prognostic Modeling for 2-Year Head and Neck Cancer-Specific Survival Using Electronic Health Record Data: A Multisite Study. *JCO Clinical Cancer Informatics*. 2023;7(7):e2200128. doi: 10.1200/cci.22.00128
- Tseng YJ, Wang YC, Hsueh PC, Wu CC. Development and validation of machine learning-based risk prediction models of oral squamous cell carcinoma using salivary autoantibody biomarkers. *BMC Oral Health*. 2022;22(1):534. doi: 10.1186/s12903-022-02607-2
- Moharrami M, Azimian Zavareh P, Watson E, Singhal S, Johnson AEW, Hosni A, et al. Prognosing post-treatment outcomes of head and neck cancer using structured data and machine learning: A systematic review. *PLoS ONE*. 2024;19(7):e0307531. doi: 10.1371/journal.pone.0307531
- Zhong MY, Zou XC, Chen PH. Explainable machine learning predicts overall survival in female bladder cancer patients after radical cystectomy. *Discov Onc*. 2025;17(1):175. doi: 10.1007/s12672-025-04275-4
- Qin Y, Li X, Sun H, Zhao W, Zhu L, Wang J, et al. Machine learning prognosis model for locally recurrent rectal cancer patients after radioactive 125I seed implantation. *Sci Rep*. 2025;16(1):2679. doi: 10.1038/s41598-025-32579-6
- Sasahira T, Kirita T. Hallmarks of Cancer-Related Newly Prognostic Factors of Oral Squamous Cell Carcinoma. *IJMS*. 2018;19(8):2413. doi: 10.3390/ijms19082413
- Brockmeyer P. Prognostic biomarkers in oral squamous cell carcinoma: current evidence and future directions. *Front Oncol*. 2025;15:1692294. doi: 10.3389/fonc.2025.1692294
- Lee MK, Joo H, Bae M, Lee Y, Noh JK, Lee YC, et al. Keratinization-related gene signature predicting survival and response to radiation in patients with HPV-negative head and neck squamous cell carcinoma via regulation of cornification and integrin signaling. *Cell Mol Biol Lett*. 2026;31(1):26. doi: 10.1186/s11658-025-00855-y
- Kou W, Guo R, Zhang Y, Xu C, Sha W, Chen Y, et al. An Eight-Gene Metabolic Prognostic Signature in Head and Neck Squamous Cell Carcinoma: Cross-Cohort Evaluation, Single-Cell Analysis, and Preliminary Functional Evaluation of PYGL and G6PD. *J Cancer*. 2026;17(11):1895-1914. doi: 10.7150/jca.134402
- Albuquerque LDA, Oliveira LQRD, Moreira DGL, França GMD, Freitas RDA, Silva JRV, et al. Multi-Omics Biomarkers in Head and Neck Squamous Cell Carcinoma: Biological Insights and Clinical Translation: A Narrative Review. *Biology*. 2026;15(15):1321. doi: 10.3390/biology15151321
- Ashwini R, Narayan M, Rajkumar K. Diagnostic and prognostic markers of oral squamous cell carcinoma- a detailed review. *Oral Oncology Reports*. 2024;10:100321. doi: 10.1016/j.oor.2024.100321
- Alsubaie NE, El Guma F, Musa AGM, Abdali HA, Awadalla M, Alosaimi B. Comparative evaluation of Cox regression and machine learning approaches for survival prediction in stage IV lung cancer. *Discov Computing*. 2026;29(1):102. doi: 10.1007/s10791-026-10014-2
- Ganapathy S, Harichandrakumar KT, Penumadu P, Tamilarasu K, Nair NS. Comparison of Bayesian, Frequentist and Machine learning models for predicting the two-year mortality of patients diagnosed with squamous cell carcinoma of the oral cavity. *Clinical Epidemiology and Global Health*. 2022;17:101145. doi: 10.1016/j.cegh.2022.101145
- Sakai T, Omura G, Eguchi K, Sakai A, Matsumoto Y, Fushimi C, et al. Prognostic prediction using recursive partitioning analysis of patients undergoing salvage surgery for locally recurrent oral squamous cell carcinoma. *Int J Clin Oncol*. 2025;30(6):1118-1126. doi: 10.1007/s10147-025-02739-9
- Prasse G, Glaas A, Meyer HJ, Zebralla V, Dietz A, Hering K, et al. A Radiomics-Based Machine Learning Perspective on the Parotid Gland as a Potential Surrogate Marker for HPV in Oropharyngeal Cancer. *Cancers*. 2023;15(22):5425. doi: 10.3390/cancers15225425
- Zhou DS, Lu YQ. Identification of a 32-gene signature that determines HPV status in head and neck cancer. *Discov Onc*. 2026;17(1):419. doi: 10.1007/s12672-026-04640-x
- Zhao D, Huang Z, Xu J, Zhang Z, Hess J, Feng B. MGP_HNSC32: A H&E Morphology-Derived 32-Gene Prognostic Signature in Head and Neck Squamous Cell Carcinoma. *Cancer Medicine*. 2026;15(11):e72427. doi: 10.1002/cam4.72427
- Gao SY, Hughes JM, Nguyen SA, McCaulay BS, Newman JG. Performance of Machine Learning Models for Prognosis Prediction in Oral Cavity Squamous Cell Carcinoma: A Systematic Review. *Cancers*. 2026;18(14):2261. doi: 10.3390/cancers18142261
- Hughes JM, Gao SY, Nguyen SA, Newman JG. Performance of Machine Learning Models for Predicting Occult Nodal Metastasis in Oral Cavity Squamous Cell Carcinoma: A Systematic Review and Diagnostic Accuracy Meta-Analysis. *Cancers*. 2026;18(14):2271. doi: 10.3390/cancers18142271
- Valero C. A novel nomogram-based proposal for prognostic staging of patients with oral cavity squamous cell carcinoma. *British Journal of Oral and Maxillofacial Surgery*. 2025;63(10):e61. doi: 10.1016/j.bjoms.2025.09.111

25. Alabi RO, Bazina M, Guntinas-Lichius O, Elmusrati M, Almangush A, Tiblom Ehrsson Y, et al. Early and late mortality among patients with T1–T3 head and neck squamous cell carcinoma: a machine learning analysis using a SEER population-based cohort study. *Sci Rep.* 2026;16(1):27640. doi: 10.1038/s41598-026-55741-0
26. Bur AM, Holcomb A, Goodwin S, Woodroof J, Karadaghy O, Shnayder Y, et al. Machine learning to predict occult nodal metastasis in early oral squamous cell carcinoma. *Oral Oncology.* 2019;92:20-25. doi: 10.1016/j.oraloncology.2019.03.011
27. Chaudhry S, Ehtesham Z, AL- Ayiedh F, Alqurouni BH, Almutairi FD. Clinicopathological Machine Learning Models for Mortality Risk Prediction in Head and Neck Squamous Cell Carcinoma: A Retrospective Analysis of the The Cancer Genome Atlas Head-Neck Squamous Cell Carcinoma (TCGA-HNSC) Cohort. *Cureus.* 2026;18(9):e116022. doi: 10.7759/cureus.116022
28. Wang P, Tian J, Zheng G, Chen Y, Ren X. Application of artificial intelligence in head and neck squamous cell carcinoma. *Annals of Medicine.* 2026;58(1):2620191. doi: 10.1080/07853890.2026.2620191
29. Lin Y, Tang M, Liu Y, Jiang M, He S, Zeng D, et al. A narrative review on machine learning in diagnosis and prognosis prediction for tongue squamous cell carcinoma. *Transl Cancer Res.* 2022;11(12):4409-4415. doi: 10.21037/tcr-22-1669
30. Cavalieri S, De Cecco L, Brakenhoff RH, Serafini MS, Canevari S, Rossi S, et al. Development of a multiomics database for personalized prognostic forecasting in head and neck cancer: The Big Data to Decide EU Project. *Head & Neck.* 2021;43(2):601-612. doi: 10.1002/hed.26515

FUNDING

No financing

CONFLICT OF INTEREST

None

AUTHORSHIP CONTRIBUTIONS

Drafting – original draft: Arlin Castro Diaz, Rigel Ramirez Fritze.
Writing–review and editing: Arlin Castro Diaz, Rigel Ramirez Fritze.