

Summary

We would like to express our sincere gratitude to Reviewer 1 for their thorough, meticulous, and highly constructive evaluation of our manuscript. The reviewer's insightful feedback has been instrumental in significantly improving the methodological transparency, narrative clarity, and overall scientific calibration of our review.

In accordance with your valuable suggestions, we have made substantial revisions throughout the manuscript. The major improvements include: * **Strict Calibration of Clinical Claims:** We have carefully tempered our language across the Abstract, Introduction, and Discussion to accurately reflect the predominantly prognostic nature of the current evidence base, explicitly labeling extrapolations regarding anti-tumoral response or adjuvant decision-making as "speculative future hypotheses". * **Methodological Clarifications:** We have refined our eligibility criteria descriptions, explicitly detailed our dual QUIPS/PROBAST approach for hybrid studies, and corrected all numerical inconsistencies within the PRISMA flow diagram. * **Refined Evidence Tiering and Meta-analytical Caveats:** We implemented a strict narrative separation ("signposting") to clearly demarcate the evidentiary strength of different biomarkers. Furthermore, we thoroughly recalibrated our discussion of the CA9 exploratory pooling, explicitly highlighting its limitations regarding the broader reproducibility challenges in the field. **Comprehensive Language Editing:** We have conducted a thorough round of English language and technical editing across the entire manuscript to correct grammatical inconsistencies, refine sentence structures, and improve overall readability and flow.

We believe these targeted revisions have resulted in a much stronger, critically balanced, and rigorously calibrated manuscript. Please find our detailed, point-by-point responses to each of your specific comments below.

Point-by-point responses

Abstract:

Comments 1:

[– The statement that these biomarkers may be relevant to “anti-tumoral response, treatment stratification, and clinically relevant decisions” feels too ambitious given that the included studies are mainly prognostic, heterogeneous, and not specifically designed to evaluate treatment response or therapeutic selection.]

Response 1:

We sincerely thank the reviewer for this insightful and methodologically astute observation. We completely agree that the included studies are fundamentally prognostic and were not specifically designed as predictive biomarker trials to evaluate direct treatment response or therapeutic selection. Framing these markers as ready for "clinically relevant decisions" was indeed too ambitious.

To address this, we have carefully revised the Abstract to ensure we do not overstate the current evidence. We now explicitly state that the included data is fundamentally prognostic. We have strictly reframed the concepts of "anti-tumoral response" and "clinical decision-making" as translational hypotheses for future research, rather than direct conclusions derived from the current datasets. By making this distinction, we

hypothesize that markers driving aggressive prognostic biology might eventually inform treatment stratification, while fully acknowledging the current limitations of the evidence base.

These revisions are provided in the Abstract, sentences 21-25 of the revised manuscript

Introduction:

Comments 2:

[- The manuscript currently states that IHC-based studies were excluded because IHC is "inherently prone to bias due to its hypothesis-driven nature," which is too categorical and methodologically debatable. The justification should be reframed more rigorously around review scope rather than implied inferiority of non-omics evidence. In addition, the Introduction repeatedly links prognostic biomarkers to cN0 neck management, adjuvant therapy selection, and surveillance intensity, but this translational framing is more aspirational than evidence-based for the included dataset.]

Response 2:

We thank the reviewer for this nuanced and constructive feedback.

First, we fully agree that labeling IHC as "inherently prone to bias" was too categorical. We have revised the "Eligibility Criteria" section to reframe this exclusion strictly around the review's intended scope: focusing exclusively on unbiased, high-throughput omics discovery workflows as opposed to *a priori* targeted IHC studies, without implying methodological inferiority.

These revisions are provided in section 2.2 (Eligibility Criteria), sentences 169-173 of the revised manuscript

Second, regarding the translational framing, we agree that directly predicting "anti-tumoral response" from purely prognostic datasets is currently aspirational. However, our study did systematically map and evaluate the included biomarkers against specific clinical scenarios (such as cN0 neck management and adjuvant therapy stratification), based on their prognostic associations (e.g., markers significantly linked to occult nodal metastasis or aggressive recurrence).

To make this distinction clear and scientifically rigorous, we have revised the final paragraph of the Introduction. We now explicitly separate the two concepts: we maintain our focus on evaluating the markers' potential for risk stratification and clinical decision-making (which our data mapping supports), while strictly isolating "anti-tumoral response" as a "hypothesized potential" where direct prediction remains an "aspirational goal requiring prospective validation." We believe this accurately reflects our analytical efforts while fully respecting the limitations of the current evidence base.

These revisions are provided in the last paragraph of the Introduction section, sentences 116-122 of the revised manuscript

Materials and Methods:

Comments 3:

[- The eligibility criteria need clarification. The population is described as “primarily surgically treated TSCC,” yet it is not clear how strictly this was applied across included studies or whether mixed-treatment cohorts were excluded if prognostic analyses remained eligible.]

Response 3:

We thank the reviewer for pointing out this ambiguity. By "primarily surgically treated," we specifically meant patients who underwent **upfront surgical resection**. We strictly excluded cohorts that received induction (neoadjuvant) therapies prior to tissue collection. While patients may have subsequently received adjuvant treatments, all analyzed molecular material across the included studies was derived exclusively from treatment-naïve specimens. We have revised Section 2.2 to explicitly clarify this criterion, sentences 136-141 of the revised manuscript

Comments 4,5:

[- The exclusion of non-English studies introduces language bias and should be acknowledged more clearly as a limitation rather than a neutral procedural choice.]

[- The decision not to search grey literature is understandable, but it should likewise be discussed as a possible source of publication bias.]

Response 4, 5:

We thank the reviewer for highlighting these important methodological constraints. We fully agree that these procedural choices carry inherent risks of language and publication bias, respectively. To provide a more transparent and balanced appraisal of our methodology, we have explicitly acknowledged both the language bias and the potential for publication bias, due to the exclusion of grey literature, within the Limitations paragraph of the Discussion section, sentences 942-948 of the revised manuscript

Comments 6:

[- The use of QUIPS for single prognostic factor studies and PROBAST for prognostic models is reasonable, but the manuscript should explain more explicitly how hybrid studies were handled, especially papers that mix single-marker findings, signatures, and nomograms.]

Response 6:

We appreciate the reviewer pointing out this important methodological detail. For "hybrid" studies presenting both biosignatures findings and composite models, such as risk scores or nomograms, we triaged the risk-of-bias assessment based on the primary prognostic output that was extracted and evaluated in our synthesis. If a study's central translational product was a predictive model or a multi-gene signature, it was appraised using PROBAST to rigorously capture risks associated with model development, parameter tuning, and potential overfitting. If the study's primary focus was the validation of individual, standalone prognostic factors, QUIPS was applied. We have added a specific paragraph to Section 2.6 (Risk of Bias and Reporting Quality Assessment) to

explicitly describe this triage methodology and ensure full transparency, sentences 237-241 of the revised manuscript

Comments 7:

[- The exploratory meta-analysis of CA9 needs more caution. The manuscript itself acknowledges that most biomarkers were too heterogeneous for pooling, and only CA9 met the criteria. However, with only two studies, moderate heterogeneity, different platforms, and likely substantial clinical and analytical differences, the practical value of pooled estimates is very limited.]

Response 7:

We completely agree with the reviewer's cautious perspective. While we initially included the exploratory pooling to quantitatively visualize the only replicated marker in our dataset, we fully acknowledge that combining just two studies with differing platforms and clinical characteristics yields an estimate of very limited practical value.

To prevent any over-interpretation of this finding, we have significantly strengthened the caveats in both the Results (Section 3.5, sentences 647-651) and Discussion (Section 4, sentences 845-849) sections. We now explicitly state that the pooled effect size is strictly illustrative, possesses highly limited practical utility, and must not be interpreted as a definitive clinical metric, but merely as a preliminary signal requiring standardized validation.

Comments 8, 9:

[- Include as supplementary material a list of the pre-selected studies that were excluded in the final screening and state the reason for the exclusion of each study.]

Results

[- The study selection reporting appears internally inconsistent. In the narrative, the title/abstract screening "reduced the set to 47 articles," and "30 were excluded," leading to 17 included studies. However, the PRISMA figure on page 9 shows a different flow with 1006 records screened, 115 reports sought, 114 reports assessed, and exclusions of 67 non-omics studies plus 31 wrong-study-design reports before reaching 17 included studies. These numbers do not reconcile cleanly and must be corrected.]

Response 8, 9:

We sincerely thank the reviewer for their meticulous reading and for catching this inconsistency between the narrative text and the PRISMA flow diagram.

The discrepancy was caused by two formatting errors in our initial diagram. First, 67 "non-omic" studies were actually excluded immediately during the initial title/abstract screening phase in Rayyan, but were mistakenly placed in the bottom full-text exclusion box in the initial PRISMA draft. Second, there was a minor typo regarding the full-text exclusions, which was mistakenly written as 31 instead of the correct 30.

We have corrected the PRISMA flow diagram (Figure 2, s.384) and thoroughly updated the narrative in Section 3.1., sentences 356-363. The numbers now perfectly reconcile: out of 1,006 screened records, 958 were excluded at the title/abstract stage (including

the non-omic studies). Consequently, 48 reports were sought for retrieval, 1 could not be retrieved, leaving exactly 47 reports assessed for full-text eligibility. Of these, 30 were excluded for ineligible study design, resulting in the final 17 included studies.

Furthermore, in accordance with the reviewer's excellent suggestion, we have compiled the complete list of these 30 excluded full-text reports. This list is now provided as **Supplementary Table S6**, detailing the specific reason for exclusion for each individual study.

Comments 10:

[- There is also a table-numbering inconsistency. Table 1 is introduced as the PICO framework, yet in the Results the authors write that "Details of the 17 eligible studies are summarized in Table 1," whereas the eligible studies are actually listed later in Table 2. This should be corrected throughout.]

Response 10:

We thank the reviewer for catching this oversight. We have corrected the table numbering throughout the manuscript; the characteristics of the eligible studies are now accurately referenced as Table 2 (sentence 387) in the Results section, and all subsequent tables follow sequentially.

Comments 11:

[- The biomarker synthesis is detailed, but the Results would benefit from stricter separation between descriptive findings and higher-confidence findings. At present, single-study associations, internally validated signatures, and replicated markers are discussed in a fairly continuous narrative, which may blur evidentiary strength.]

Response 11:

We thank the reviewer for this excellent point. We fully agree that the evidentiary strength of the various biomarkers needed clearer demarcation to prevent the blending of exploratory and higher-confidence findings.

The current narrative structure of Sections 3.3.1 and 3.3.2 was strictly designed around the biological and methodological comparisons performed in the primary studies (i.e., tumor-to-tumor comparisons in 3.3.1, versus tumor-to-normal comparisons in 3.3.2). To preserve this important biological context while fulfilling the reviewer's request for stricter separation of evidentiary strength, we implemented a robust "signposting" strategy.

First, we added a new introductory paragraph at the beginning of Section 3.3. (sentences 416-425) This paragraph explicitly explains the biological split between the sub-sections and prominently directs the reader to **Section 3.6**, where our formal Evidence Tiering is comprehensively detailed. Second, we have revised the narrative within Sections 3.3.1 and 3.3.2, introducing explicit qualifier phrases to immediately signal the evidentiary weight of the discussed markers. We believe this dual approach provides the strict separation requested without dismantling the biological logic of the synthesis.

Comments 12:

[- The statement that Figure 4 shows “the most reliable prognostic biomarkers” is too strong. Reliability is not demonstrated here in a formal way; at best, the figure shows the most statistically robust biomarkers among those with extractable multivariable OS estimates. The wording should be revised. Similarly, the CA9 section should not imply that exploratory pooling meaningfully resolves the broader reproducibility problem.]

Response 12:

We completely agree with the reviewer’s assessment. The term “reliable” was indeed an overstatement. We have carefully revised the text and the legend for Figure 4 to state exactly what the data supports, using the reviewer’s excellent phrasing: *“the most statistically robust biomarkers among those with extractable multivariable OS estimates.”* – (sentence 653-654)

Regarding the CA9 meta-analysis, we fully agree that pooling does not cure the underlying heterogeneity. Building upon our revisions for Comment 7, we have further refined the Discussion to explicitly state that this exploratory pooling *“does not resolve the broader reproducibility challenges inherent in this field”* and serves strictly as an illustrative signal rather than a definitive clinical metric. (s.848-849)

Comments 13:

[- The REMARK assessment is also presented somewhat optimistically. The manuscript says reporting quality was generally acceptable and that most studies scored above 80%, yet nearly all studies remained moderate- or high-risk by QUIPS/PROBAST. That contrast is actually important and should be made more explicit in the Results rather than only later in the Discussion.]

Response 13:

We strongly agree with the reviewer’s insightful observation. The distinction between reporting quality (REMARK) and true methodological rigor or risk of bias (QUIPS/PROBAST) is indeed a crucial finding of our review, and it was previously understated in the Results section.

We have revised the Quality Assessment narrative in the Results (Section 3.4) to make this contrast explicit. We added a dedicated paragraph (s. 607-613) explicitly stating that while the *reporting* of the studies was generally adequate (>80% REMARK scores), the underlying *methodologies* and statistical designs remained heavily flawed, as evidenced by the moderate-to-high risk of bias in nearly all studies. We believe bringing this stark contrast forward into the Results significantly strengthens the manuscript’s critical appraisal of the current literature.

Discussion:

Comments 14:

[- The manuscript includes what appears to be an editorial/template instruction embedded in the body text: “Authors should discuss the results and how they can be interpreted...” This must be removed.]

Response 14:

We apologize for this oversight and thank the reviewer for catching this residual journal template text. The instructional sentence has been entirely removed from the Discussion section. (s. 760)

Comments 15:

[- The discussion of biomarkers as possible criteria for anti-tumoral response, treatment sensitivity, adjuvant decision-making, and cN0 management is speculative. The authors do acknowledge this later, but the framing remains broader than the actual evidence base, which is mostly prognostic and rarely treatment-annotated. That interpretation should be more explicitly labeled as a future hypothesis.]

Response 15:

We appreciate the reviewer's cautious perspective and completely agree. Our initial framing was indeed too broad and confident given the strictly prognostic nature of the current evidence base.

To address this, we have completely rewritten the relevant paragraph in the Discussion. We removed the overly optimistic phrasing and explicitly stated right from the opening sentence that the current evidence base is strictly prognostic. Furthermore, we now explicitly label any potential extension of these markers into predictive roles (such as treatment sensitivity or adjuvant decision-making) as a "*speculative future hypothesis*." Finally, we emphasized that establishing true clinical relevance for these applications will strictly require validation in prospective, *treatment-annotated cohorts*. (s. 881-921)

Comments 16:

[- The CA9 discussion is biologically plausible and probably the most convincing part of the manuscript, but it still needs tighter calibration.]

Response 16:

We are highly encouraged that the reviewer found the biological rationale for CA9 to be convincing. We completely agree, however, that the initial phrasing regarding its recurrence in our dataset required tighter calibration to avoid overstating the strength of this finding.

We have carefully revised the opening sentences of the CA9 discussion. We removed the term "internal replication," replacing it with the more accurate description of "a rare instance of marker overlap." Furthermore, we explicitly recalibrated the interpretation of this recurrence, noting that it serves as an "*encouraging, albeit preliminary, signal rather than definitive proof of broad reproducibility*." We believe these targeted adjustments, combined with the strict caveats regarding the meta-analysis added in response to Comments 7 and 12, provide the precise calibration the reviewer requested. (s. 823-826)

Conclusion:

Comments 17:

[- The Conclusion is better balanced than the abstract and appropriately states that the evidence is hypothesis-generating rather than practice-changing. That said, it would be improved by one final sentence explicitly stating that no biomarker is currently ready for routine prognostic use in TSCC. This would align the conclusion with the risk-of-bias findings and avoid any residual ambiguity.]

Response 17:

We thank the reviewer for this excellent suggestion, which provides the perfect, unambiguous closing to our manuscript. We are glad the reviewer found the overall conclusion to be well-balanced.

To eliminate any residual ambiguity and to perfectly align the conclusion with our QUIPS/PROBAST risk-of-bias findings, we have added the requested explicit statement to the very end of the Conclusion section. The manuscript now concludes with the following sentence: *"Ultimately, given the pervasive methodological heterogeneity and the moderate-to-high risk of bias across the current literature, no single biomarker or multi-gene signature identified in this review is currently ready for routine clinical or prognostic use in TSCC. (s. 989-992)*

Comments on the Quality of English Language:

[The manuscript needs another careful round of English and technical editing. There are multiple small but noticeable issues.]

Response on the Quality of English Language:

We appreciate the reviewer pointing this out and apologize for the grammatical and structural inconsistencies in our original submission. In response, the entire manuscript has undergone a comprehensive and careful round of English language and technical editing. We have thoroughly proofread the text to correct grammatical errors, refine sentence structures, remove redundancies, and ensure the consistent and accurate use of technical terminology. We believe the readability, flow, and academic tone of the revised manuscript have been significantly improved.