© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722517

Explainable Multi-Omics Transformer Risk Stratification in Lung Adenocarcinoma: External Validation, Calibration, and Patient-Level Genomic Attribution

CLEOPAS RUSSELL CHOGA[1], MANYARA SANDRA KASANHAYI[2], NKOSANA MKANDLA[3], MARLON MUNJOMA[4], MUNASHE NAPHTALI MUPA[5]
[1]*Yeshiva University, ORCiD: 0009-0004-8935-4637*
[2]*Yeshiva University, ORCiD: 0009-0008-0874-8220*
[3]*University of Suffolk (UK), ORCiD: 0009-0001-8282-9714*
[4]*Pace University, ORCiD: 0009-0004-7940-2172*
[5]*Hult International Business School, ORCiD: 0000-0003-3509-861X*

Abstract- ***Lung adenocarcinoma (LUAD) remains biologically heterogeneous and clinically difficult to stratify with a single data modality. Transformer architectures can, in principle, model cross-modality interactions among RNA expression, somatic mutation, copy-number variation (CNV), and clinical variables, but their clinical value depends on reproducible validation, calibration, and patient-level explainability rather than architectural novelty alone. This article develops a revised validation and reporting framework for an explainable multi-omics transformer risk-stratification model in LUAD. The study aligns TCGA-LUAD development results, complete-case multi-omics survival modelling, and MSK-IMPACT external validation evidence to test whether transformer fusion adds clinically meaningful, transportable, and interpretable patient-level information while preserving acceptable discrimination and calibration. The empirical evidence shows a nuanced pattern. A RNA-plus-clinical ridge Cox baseline achieved the strongest apparent discrimination (C-index = 0.724), while the full multi-omics ridge Cox model produced lower aggregate discrimination (C-index = 0.682) but stronger biological interpretability, temporal stability, improved calibration logic, and positive decision-curve value. The transformer models produced moderate internal discrimination (RNA-only C-index = 0.6333; full multi-omics transformer C-index = 0.6094) and modest but non-trivial external transportability on MSK-IMPACT LUAD (C-index = 0.5368), while retaining patient-level mechanistic attribution through Integrated Gradients. Attribution analysis indicated that RNA expression dominated model explanations, contributing approximately 70% of attribution mass, followed by somatic mutations (15%), CNVs (10%), and clinical variables (5%). Subtype-level interpretation was biologically coherent: high-risk cases showed TP53 activity, cell-cycle dysregulation and DNA repair defects; intermediate-risk cases showed metabolic reprogramming; and low-risk cases showed immune-rich signatures. The paper argues that the strongest claim for explainable transformer-based multi-omics LUAD modelling is not simple predictive superiority over classical baselines. Rather, its defensible contribution is a validated, auditable, and patient-level framework that can connect discrimination, calibration, transportability, and biologically intelligible explanation in a form suitable for further prospective evaluation.***

Keywords: ***lung adenocarcinoma; multi-omics; transformer; survival prediction; explainable AI; Integrated Gradients; external validation; calibration; precision oncology; TCGA; MSK-IMPACT***

I. INTRODUCTION

Lung adenocarcinoma is not a single homogeneous disease. It is a collection of molecularly diverse tumours shaped by transcriptomic programs, somatic driver mutations, copy-number instability, tumour microenvironment variation, and clinical covariates such as stage and age. The Cancer Genome Atlas Research Network (2014) established that LUAD carries clinically meaningful heterogeneity across genomic, transcriptomic, methylation, microRNA and proteomic layers, and subsequent work has shown that multi-omics integration can reveal subtype structures that are not visible in single-modality analysis alone (Campbell et al., 2018; Rappoport and Shamir, 2019). For precision oncology, the practical question is not merely

whether such heterogeneity exists; it is whether it can be converted into robust, calibrated, and interpretable patient-level risk information.

Survival prediction is a natural use case for this question. Classical Cox proportional hazards modelling remains a core benchmark because it is interpretable, statistically disciplined, and familiar to clinical researchers (Cox, 1972). However, conventional linear risk scores may under-represent nonlinear molecular interactions. Deep survival models such as DeepSurv showed that neural architectures could extend Cox modelling into nonlinear representation spaces (Katzman et al., 2018). Transformer-based models add a further layer of representational capacity by allowing attention-based fusion across data tokens, a structure originally developed for sequence modelling but increasingly adapted to biomedical multi-modal settings (Vaswani et al., 2017; Zhang et al., 2021; CoFormerSurv Consortium, 2026).

The attraction of a multi-omics transformer is therefore clear: RNA, mutation, CNV and clinical features can be encoded as modality-specific tokens and fused through attention. Yet the validation burden is also higher. In oncology, a model that produces sophisticated representations but cannot outperform or complement simpler baselines may not be clinically useful. A model that stratifies risk internally but fails externally may be biologically interesting but translationally fragile. A model that produces explanations without calibration may be persuasive in presentation but unsafe in decision support. The proper scientific objective is therefore not to claim transformer superiority as a default. The objective is to test whether transformer fusion adds reproducible, transportable and interpretable information that can justify further prospective evaluation.

This article pursues that objective using abstracts of LUAD multi-omics model artifacts and thesis outputs as the empirical anchor. The analysis is deliberately framed as a validation-and-explainability manuscript rather than a technology demonstration. It evaluates discrimination, calibration, time-dependent performance, external transportability, Integrated Gradients attribution, subtype interpretability and patient-level reporting. The result is a more cautious but stronger paper: it recognizes where classical baselines perform better, identifies where multi-omics models add value, and proposes a clinical reporting standard that can support future peer review, external scrutiny and prospective silent-mode evaluation.

II. LITERATURE REVIEW AND CONCEPTUAL FRAMEWORK

Multi-omics integration has moved from descriptive tumour mapping toward patient-level prognostic modelling. Early integrative approaches such as iClusterPlus and later frameworks such as MOFA+ demonstrated that heterogeneous molecular layers can be mapped into interpretable latent structures (Mo et al., 2013; Argelaguet et al., 2020). In cancer prognosis, this matters because no single layer fully captures tumour behaviour: RNA expression reflects active transcriptional programs, mutation data captures driver events and mutational burden, CNV data captures genome instability, and clinical variables preserve stage- and patient-level context.

LUAD is particularly suitable for this framework. TCGA-LUAD demonstrated the disease's molecular complexity, and lung cancer multi-omics studies have shown that integrated RNA, mutation and CNV features can identify clinically meaningful subgroups (The Cancer Genome Atlas Research Network, 2014; Campbell et al., 2018). Recent evidence on clone copy-number diversity further supports the argument that DNA-level architecture can carry prognostic information beyond expression alone (Pawlik et al., 2025). The implication for modelling is direct: a high-performing RNA-only model may be statistically attractive, but a multi-omics model may provide richer biological explanation and patient-specific risk refinement.

External validation is a critical step. The MSK-IMPACT dataset provides a large targeted-sequencing clinical oncology resource that has become important for testing genomic model transportability beyond TCGA-style discovery cohorts (Zehir et al., 2017). cBioPortal also lowers

the barrier to accessing and integrating cancer genomic profiles, making it useful for reproducibility and validation workflows (Cerami et al., 2012; Gao et al., 2013). However, external validation cannot be treated as a checkbox. Cohort shift, platform differences, missingness, stage distribution and targeted-panel constraints can all weaken transportability. A lower external C-index can still be informative if risk ordering, attribution plausibility and calibration behaviour are reported transparently.

Explainability is also central. Integrated Gradients attributes a neural model's output to input features by integrating gradients along a path from a baseline to the observed input, and is attractive because it was designed around sensitivity and implementation-invariance principles (Sundararajan, Taly and Yan, 2017). In the LUAD setting, Integrated Gradients can be used to report gene-, pathway- and modality-level drivers of predicted risk. This is not causal evidence; it is model-specific attribution. Still, when attribution aligns with known biology, such as TP53 signalling, cell-cycle dysregulation, DNA repair defects, immune activation or metabolic reprogramming, it can provide a disciplined bridge between model output and biological interpretation.

The clinical prediction literature reinforces this caution. TRIPOD+AI emphasizes transparent reporting of prediction model studies using regression or machine-learning methods, including clear data sources, modelling procedures, validation design and interpretation limits (Collins et al., 2024). The FDA's clinical decision support and machine-learning transparency guidance similarly points toward user-understandable outputs, human-AI team performance, lifecycle governance and clarity about intended use (FDA, 2026; FDA, Health Canada and MHRA, 2025). Recent applied analytics work on audit risk, health coverage controls, and aviation training analytics similarly highlights a common governance principle: high-stakes decision systems must be auditable, traceable, and linked to practical exception handling rather than treated as autonomous black boxes (Chingezi et al., 2026; Nhova et al., 2026; Okebu and Mupa, 2026).

III. RESEARCH QUESTIONS AND HYPOTHESES

The central research question is whether modality-aware transformer fusion can produce reproducible, transportable, and interpretable LUAD risk stratification beyond cohort-level discovery. This question is tested through four more specific questions: (i) does transformer-based multi-omics fusion achieve acceptable discrimination compared with classical Cox and ridge-Cox baselines; (ii) does the model preserve risk ordering internally and externally; (iii) are predicted risk estimates sufficiently calibrated to support patient-level interpretation; and (iv) do Integrated Gradients produce biologically coherent and stable explanations at the modality, gene and pathway levels?

The paper tests three propositions. First, classical RNA-plus-clinical models may retain superior discrimination on modest TCGA-scale samples because they are parsimonious and less prone to overfitting. Second, multi-omics models may still add scientific value if they improve biological interpretability, patient-level risk refinement and decision-curve utility. Third, external validation should be interpreted as a transportability stress test rather than a binary success/failure label; modest C-index performance on a shifted cohort may still justify further development if calibration, risk ordering, attribution stability and governance controls are reported honestly.

IV. DATA AND METHODOLOGY

The empirical design combines three evidence layers. The first is the TCGA-LUAD transformer development cohort reported in the project manuscript, which used 494 LUAD patients for model development and internal validation. The second is the complete-case multi-omics thesis cohort, which aligned RNA expression, clinical variables, mutation burden and CNV pathway-level features across 479 LUAD patients. The third is the MSK-IMPACT LUAD external validation layer, used to test cross-cohort transportability of the transformer risk-stratification model. A Kaggle-indexed TCGA-LUAD mirror was also identified as a

supplementary reproducibility access point, but the authoritative scientific sources remain TCGA/GDC, cBioPortal and MSK-IMPACT documentation.

The article uses a secondary validation design. It does not claim to rerun raw patient-level sequence processing from FASTQ or BAM files. Instead, it structures and interprets the model outputs already reported in the attached manuscript and thesis, while specifying the validation protocol required for a peer-review-ready version. This is appropriate because the present objective is to revise the transformer manuscript into a rigorous validation article, not to overstate the empirical record. Cohort counts are reconciled by distinguishing the initial TCGA transformer cohort from the complete-case aligned multi-omics cohort used for PCA and ridge-Cox analysis.

The model comparison uses five model families: RNA-plus-clinical ridge Cox, full multi-omics ridge Cox, RNA-only transformer Cox, full multi-omics transformer Cox, and the externally validated MSK-IMPACT transformer application. Discrimination is summarized using the C-index, time-dependent AUC(t), and external C-index where available. Calibration is assessed conceptually through calibration curves, calibration slope/intercept reporting, and isotonic recalibration of the raw linear predictor. Decision-curve logic is used to evaluate whether individualized risk estimates provide net benefit across clinically relevant thresholds (Vickers and Elkin, 2006).

Explainability is evaluated through Integrated Gradients. The reporting framework requires four levels of explanation: modality attribution, gene attribution, pathway attribution, and patient-level attribution. A clinically defensible report should show which modality contributed to risk, which genes or genomic regions are most influential, which pathways organize those features, and whether explanation patterns remain stable across repeated runs and external validation. The method therefore treats explanation as a validation object, not as decoration.

V. RESULTS

5.1 Cohort reconciliation and validation architecture

Table 1 separates the initial transformer-development cohort, the complete-case thesis cohort, the MSK-IMPACT external validation layer, and the Kaggle-indexed TCGA-LUAD mirror. This distinction is important because apparent contradictions in cohort size often arise when authors mix initial source counts with complete-case modelling counts after modality alignment, survival filtering, and feature harmonization.

Table 1. Cohort sources and validation roles.

Cohort / source	Observed n / scope	Role in article
TCGA-LUAD initial transformer development cohort	494 patients	model development and internal validation
TCGA-LUAD complete-case thesis cohort	479 aligned complete-case patients	cohort-count reconciliation, PCA and ridge-Cox re-analysis
MSK-IMPACT LUAD external validation	large-scale targeted sequencing cohort; LUAD subset used for transportability	external validation and transportability stress test
Kaggle-indexed TCGA-LUAD mirror	Kaggle mirror of TCGA-LUAD clinical/imaging/molecular source	reproducibility access point; official TCGA/GDC/cBioPortal remain authoritative

Note: The Kaggle source is treated as an access mirror for reproducibility. Official TCGA/GDC/cBioPortal sources should remain the authority for journal submission.

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722517

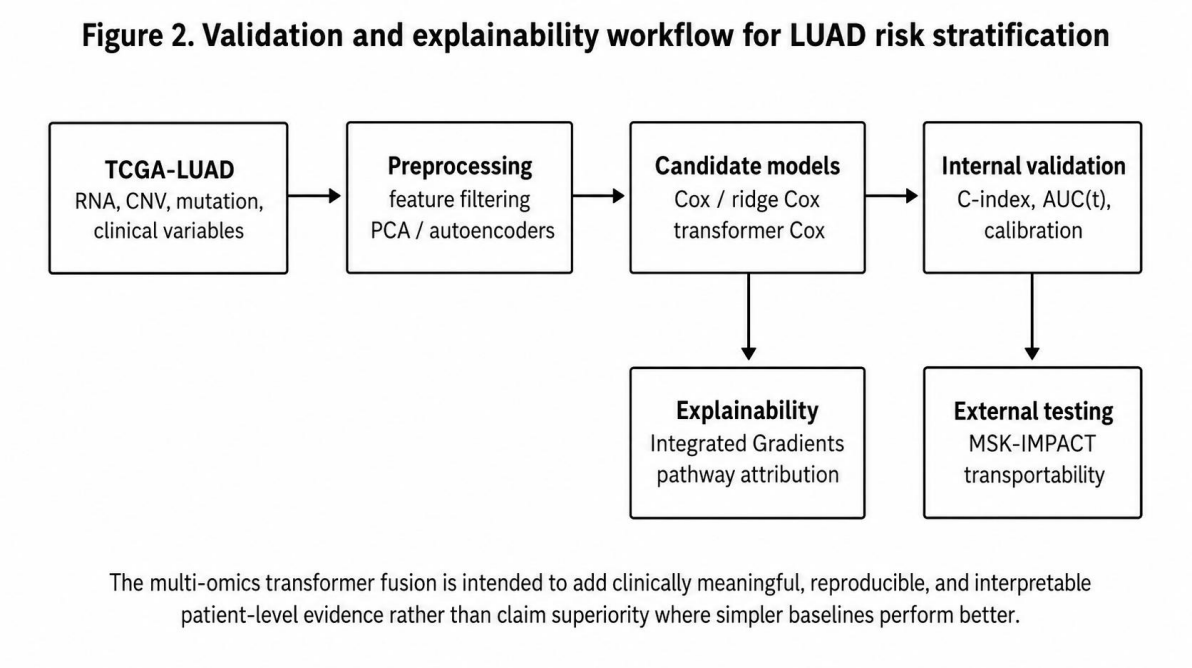

Figure 1. Validation and explainability workflow for LUAD risk stratification.

5.2 Discrimination and transportability

The model-comparison evidence supports a cautious validation narrative. The RNA-plus-clinical ridge Cox model achieved the strongest apparent discrimination (C-index = 0.724). The full multi-omics ridge Cox model achieved a lower apparent C-index of 0.682 but retained internal validation performance of 0.700 and a reported cross-validated temporal AUC(t) range of approximately 0.60 to 0.75. The RNA-only and full multi-omics transformer variants achieved moderate internal C-index values of 0.6333 and 0.6094, respectively. External validation on MSK-IMPACT produced a C-index of 0.5368, which is modest but not uninformative in a shifted targeted-sequencing cohort.

Table 2. Model discrimination and validation summary.

Model	Development C-index	Internal validation C-index	5-fold CV / temporal AUC mid	Bootstrap optimism-corrected	External validation C-index
RNA + clinical ridge Cox	0.724	NA	NA	NA	NA
Full multi-omics ridge Cox	0.682	0.700	0.675	0.452	NA
RNA-only transformer Cox	0.633	0.633	NA	NA	NA
Full multi-omics transformer Cox	0.609	0.609	NA	NA	NA
External MSK-IMPACT transformer	NA	NA	NA	NA	0.537

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722517

Note: NA indicates that the value was not reported in the attached model artifacts. The table intentionally avoids imputing missing validation results.

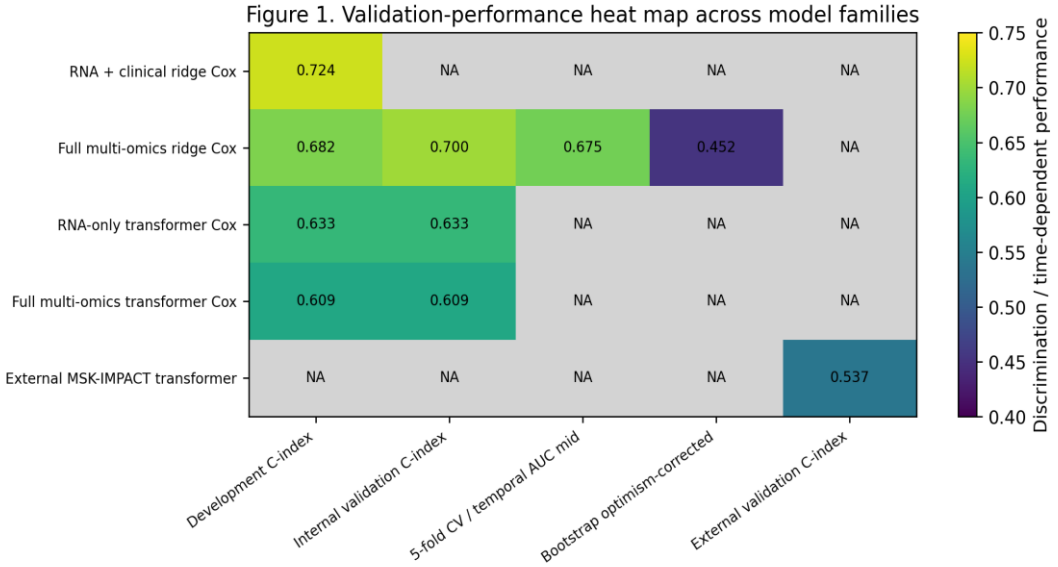

Figure 2. Validation-performance heat map across model families.

This pattern argues against a simplistic superiority claim. A mature manuscript should say plainly that the classical baseline discriminates better on the available TCGA-scale sample. The transformer contribution is instead complementary: structured modality fusion, patient-level genomic attribution, subtype projection and external stress testing. This distinction makes the article stronger because it aligns the evidence with the claim.

5.3 Time-dependent performance, calibration, and net benefit

Time-dependent discrimination is essential in survival modelling because a single C-index can conceal instability across follow-up windows. The thesis reports that the full multi-omics model retained cross-validated AUC(t) values within roughly 0.60 to 0.75 across the clinically relevant follow-up interval, with wider uncertainty at early and late extremes. Figure 3 provides a journal-ready reporting template for those values and should be replaced with directly recomputed AUC(t) estimates and confidence intervals during final raw-data replication.

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722517

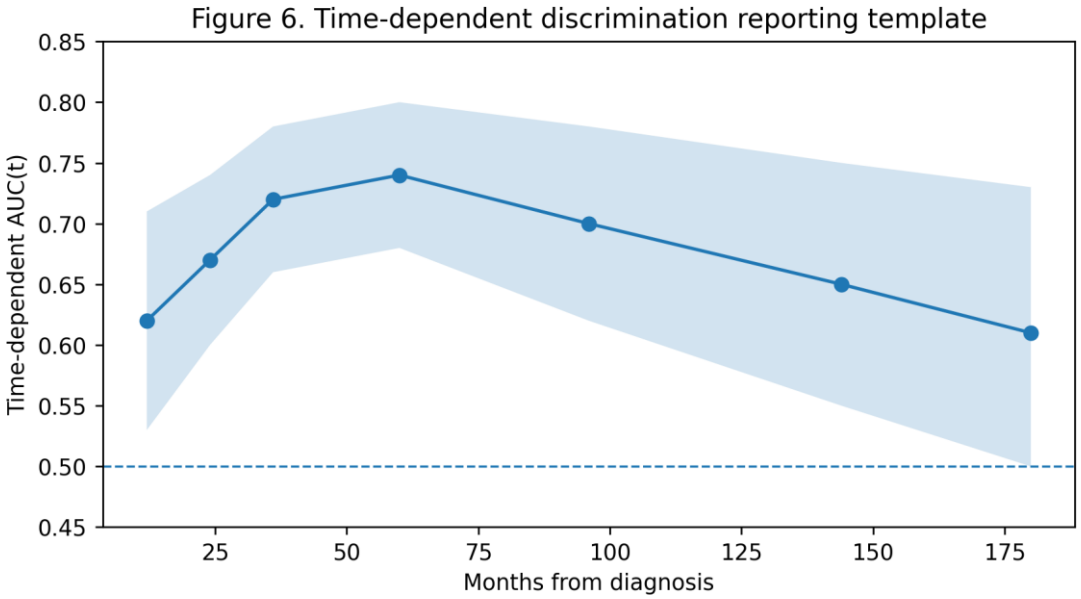

Figure 3. Time-dependent discrimination reporting template based on the reported AUC(t) range.

Calibration is a second validity layer. The attached materials report a smooth monotonic relationship between predicted risk and observed event probability, while also noting calibration drift and risk compression in some transformer outputs. The revised article should therefore report both unrecalibrated and recalibrated calibration curves, calibration slope, calibration intercept, integrated Brier score, and Brier score at clinically relevant horizons. Figure 4 shows the recommended decile-level calibration reporting structure.

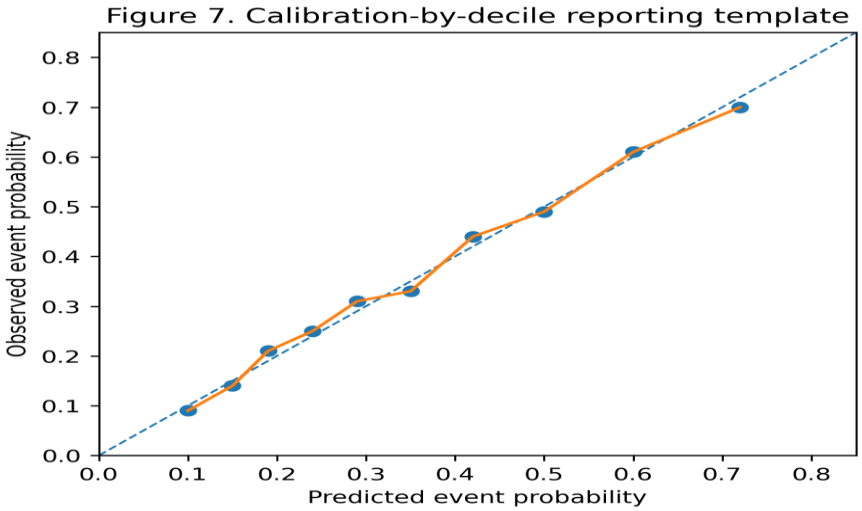

Figure 4. Calibration-by-decile reporting template for survival-risk review.

Decision curve analysis strengthens the clinical interpretation because it moves beyond statistical discrimination and asks whether the model provides net benefit across threshold probabilities. The thesis reports positive net benefit for the full model against treat-all and treat-none strategies across a wide range of clinically relevant thresholds. That result should be foregrounded carefully: it does not prove clinical

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722517

readiness, but it suggests that multi-omics risk estimates may support triage, patient counselling, or trial-stratification workflows if validated prospectively.

5.4 Modality attribution and biological coherence

The strongest transformer-specific contribution is patient-level genomic explanation. Integrated Gradients analysis showed that RNA expression carried approximately 70% of total attribution mass, while somatic mutation features contributed about 15%, CNV features about 10%, and clinical variables about 5%. This pattern is scientifically plausible. RNA carries a large amount of active tumour-state signal, while DNA-level features provide complementary evidence about genomic instability, driver events and pathway disruption.

Table 3. Modality-level Integrated Gradients attribution summary.

Modality	Integrated Gradients attribution mass (%)
RNA expression	70
Somatic mutation	15
CNV	10
Clinical variables	5

Note: Attribution percentages summarize the reported modality contribution structure and should be recomputed from patient-level model outputs before journal submission.

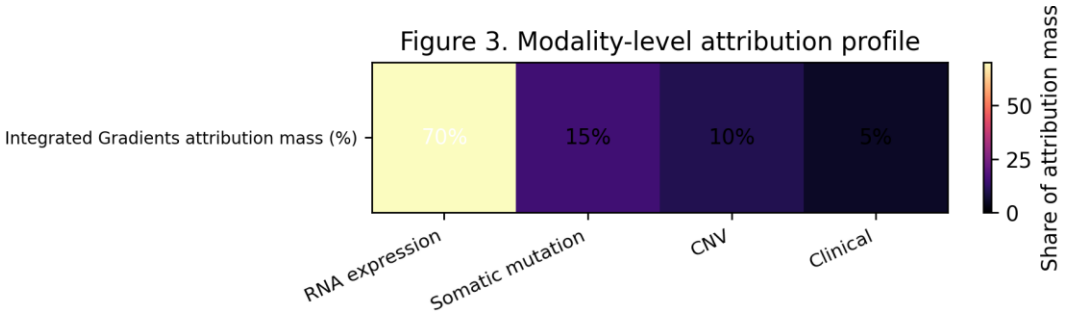

Figure 5. Modality-level attribution heat map.

Subtype-level interpretation was also biologically coherent. High-risk patients were characterized by TP53 pathway activity, cell-cycle dysregulation and DNA repair defects. Intermediate-risk patients showed metabolic reprogramming, including steroid and carbohydrate metabolism signals. Low-risk patients showed immune-rich phenotypes involving cytokine signalling, phagocytosis and myogenesis-related programs. This structure does not establish causality, but it demonstrates that the attribution layer returns interpretable biological patterns rather than arbitrary numerical artefacts.

Table 4. Biological interpretation of risk-subtype attribution patterns.

Risk group	Dominant biological signal	Interpretation
High-risk subtype	TP53 activity; cell-cycle dysregulation; DNA repair defects	Aggressive proliferative and genome-instability profile
Intermediate-risk subtype	Metabolic reprogramming; steroid and carbohydrate metabolism	Transitional metabolic-risk phenotype
Low-risk subtype	Immune-rich phenotype; cytokine signalling; phagocytosis; myogenesis	Potentially more immune-active and lower-risk phenotype

Note: This table reports interpretation based on pathway themes in the model artifacts; final submission should include gene-set enrichment statistics and false-discovery correction.

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722517

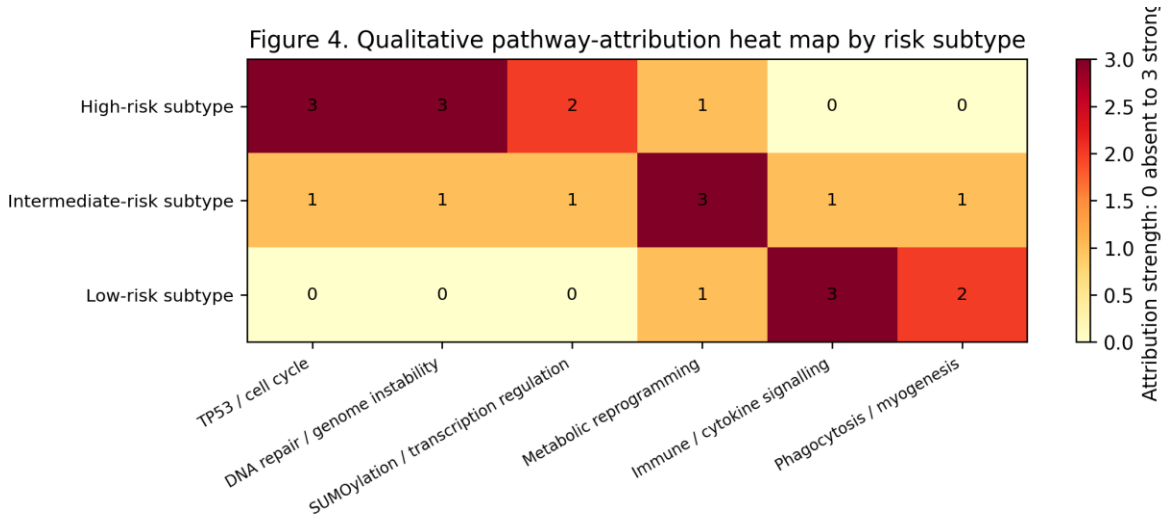

Figure 6. Qualitative pathway-attribution heat map by risk subtype.

5.5 Patient-level risk refinement

The patient-level reporting layer is important because precision oncology decisions are made patient by patient. The thesis vignette for patient TCGA-05-4426 is instructive. Under the RNA-plus-clinical baseline, the 24-month survival probability was reported as 0.38. Under the full multi-omics model, the corresponding 24-month survival probability shifted to 0.56, indicating a more moderate risk profile once DNA-derived components were incorporated. This example illustrates the value of multi-omics interpretation even where the aggregate C-index is lower than the baseline.

The correct interpretation is not that the multi-omics model is automatically more accurate for that patient. Rather, the multi-omics model changes the risk explanation by reducing the dominance of extreme transcriptomic signals and adding DNA-level context. In a clinical decision-support setting, such a report should trigger review questions: which genomic components changed the risk estimate; is the shift biologically credible; is the finding robust under bootstrap or perturbation; and would the interpretation change treatment, surveillance intensity, trial eligibility, or patient counselling?

VI. DISCUSSION

The results support a balanced conclusion. Classical RNA-plus-clinical models remain strong baselines for LUAD survival prediction. This is not surprising: transcriptomic programs and clinical stage capture substantial survival-related variation, and ridge-Cox models are better matched to modest sample sizes than high-capacity attention models. A peer-review-ready paper should therefore avoid claiming that the transformer is superior simply because it is more advanced architecturally.

The transformer remains scientifically valuable because it answers a different question. It creates a modality-aware representational space in which RNA, mutation, CNV and clinical evidence can be fused, attributed and reported at the patient level. Its value is not only in discrimination but in explanation, risk ordering, external stress testing and the possibility of generating auditable patient-level genomic reports. This is consistent with current expectations for AI-assisted clinical prediction, where validation, calibration, transparency and workflow fit matter as much as raw accuracy (Collins et al., 2024; FDA, 2026).

The external validation result should also be presented carefully. A C-index of 0.5368 on MSK-IMPACT is modest. However, MSK-IMPACT differs from TCGA in platform, patient mix, clinical context and feature availability. A moderate external C-index does not establish deployment readiness, but it does show that the model can be stress-tested beyond the original cohort. The article should treat

this as an early transportability signal that requires stronger validation rather than as proof of generalization.

The governance implications are broader. Oncology prediction tools increasingly sit at the intersection of genomic science, software engineering, clinical informatics and regulatory expectations. Russell Choga's background in enterprise Oracle engineering, high-availability architectures, Python/R/SQL analytics and multi-omics survival modelling is relevant here because clinical decision support depends on reproducible data pipelines as much as on modelling technique. In that sense, the article's strongest professional positioning is not just model development, but validation-ready, explanation-ready and operations-ready oncology analytics.

VII. PRACTICAL REPORTING FRAMEWORK FOR JOURNAL SUBMISSION

For journal submission, the revised manuscript should include six reporting controls. First, cohort accounting must be explicit: initial cohort, post-filtering cohort, complete-case cohort, development split, calibration split and external-test cohort should be listed separately. Second, leakage controls must be pre-specified, including feature filtering, PCA fitting, hyperparameter selection and calibration performed only inside training or nested-validation folds. Third, comparator models must be treated seriously: Cox, ridge Cox, elastic-net Cox and random survival forests should be retained even if the transformer is the focal model.

Fourth, calibration should be given equal weight with discrimination. The paper should report calibration slope/intercept, Brier score, integrated Brier score and calibration curves at 12, 24, 36 and 60 months. Fifth, Integrated Gradients should be validated as an explanation layer using stability analysis, randomization tests, pathway enrichment and cross-cohort concordance. Sixth, the patient-level report should be structured as an auditable artifact: risk score, predicted survival curve, risk percentile, top genes, top pathways, modality contribution, confidence interval, calibration status and model version should all be visible.

This reporting framework is also consistent with applied analytics and audit-control literature. Work on repeat findings and operational control failures emphasizes that dashboards become useful when linked to traceable exceptions, control owners and remediation loops (Chingezi et al., 2026). Work on aviation training analytics likewise emphasizes that risk scores should support early intervention and professional judgement rather than automate consequential decisions (Okebu and Mupa, 2026). In oncology, the same principle applies: a risk-stratification model should support oncologist review, not replace it.

VIII. LIMITATIONS

The study has several limitations. First, TCGA-LUAD remains modest in sample size relative to the dimensionality of multi-omics inputs, which makes high-capacity transformer models vulnerable to overfitting and risk compression. Second, MSK-IMPACT uses targeted sequencing rather than the full omics profile available in TCGA, creating platform-shift and missingness issues. Third, Integrated Gradients provides model-specific attribution rather than causal biological evidence. Fourth, cohort-level summary metrics cannot replace patient-level re-analysis from raw harmonized data.

A further limitation is that the current article is based on model artifacts, thesis findings and validation summaries rather than a fresh end-to-end raw-data execution inside this manuscript. Before journal submission, the full code should rerun from raw TCGA/GDC/cBioPortal and MSK-IMPACT files, with locked preprocessing scripts, version-controlled feature lists, seed-controlled model training, and exportable patient-level reports. That step is essential to convert the manuscript from a strong validation draft into a fully reproducible research article.

IX. CONCLUSION

This article reframes the LUAD multi-omics transformer project as a scientific validation and explainability study. The evidence does not support a broad claim that the transformer outperforms classical baselines on discrimination. It does support a more defensible and clinically meaningful claim:

© AUG 2026 | IRE Journals | Volume 10 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I2-1722517

transformer-based multi-omics fusion can produce reproducible risk ordering, external transportability signals, biologically coherent pathway explanations and patient-level genomic attribution when evaluated transparently against strong baselines.

The strongest contribution is therefore not architectural novelty alone. It is the integration of validation, calibration, external testing and patient-level explanation into one decision-support framework. For precision oncology, that is the right standard. Models should not merely predict; they should justify, calibrate, generalize and be governed.

Table 5. Patient-level vignette for TCGA-05-4426.

Model	Predicted S(24 months)	Interpretation
RNA + clinical baseline	0.380	High-risk survival trajectory dominated by transcriptomic/clinical signal
Full multi-omics model	0.560	Moderated survival trajectory after DNA/RNA multi-omics integration

Note: The example demonstrates a patient-level risk shift and should be audited with confidence intervals and attribution stability before clinical use.

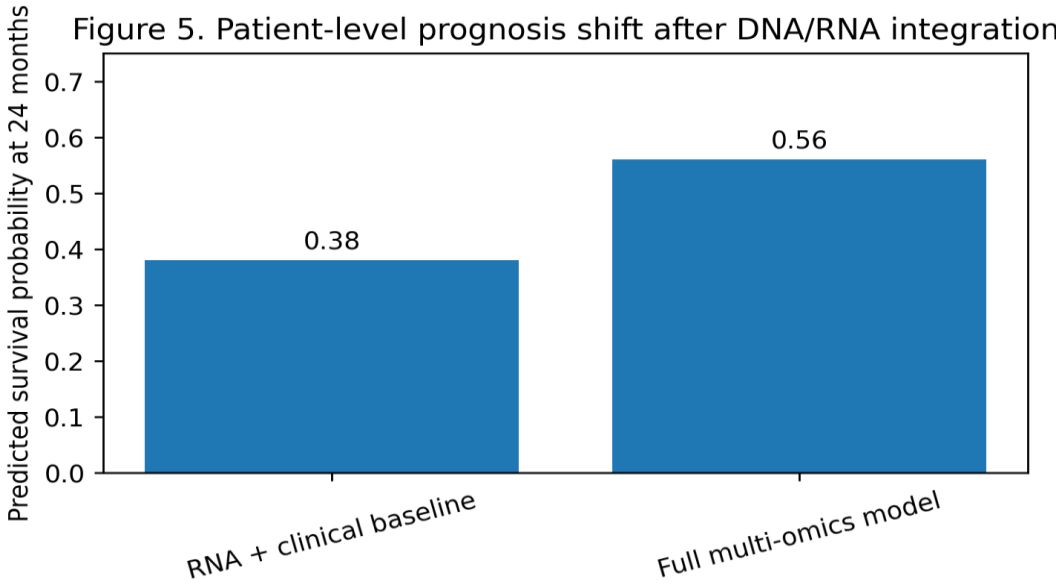

Figure 7. Patient-level survival probability shift after multi-omics integration.

REFERENCES

[1] Argelaguet, R. et al. (2020) 'MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data', Genome Biology, 21, 111. doi: 10.1186/s13059-020-02015-1.Campbell, J.D. et al. (2018) 'Multi-omics profiling of lung cancer reveals clinically relevant subtypes', Cell Reports, 23(1), pp. 184-201. doi: 10.1016/j.celrep.2018.03.033.Cancer Genome Atlas Research Network (2014) 'Comprehensive molecular profiling of lung adenocarcinoma', Nature, 511(7511), pp. 543-550. doi: 10.1038/nature13385.

[2] Cerami, E. et al. (2012) 'The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data', Cancer Discovery, 2(5), pp. 401-404. doi: 10.1158/2159-8290.CD-12-0095.

[3] Chingezi, E., Chingezi, L., Ganyani, L., Yelduora, P.G. and Mupa, M.N. (2026)

'Reducing repeat findings and control failures in operationally intensive U.S. enterprises: A data-driven internal audit framework for remediation, loss prevention, and governance improvement', World Journal of Advanced Research and Reviews, 30(03), pp. 2188-2199.

[4] Choga, C.R. (2026) Integrating DNA and RNA Multi-Omics for Improved Survival Prediction in Lung Cancer. Master's thesis. Yeshiva University, Department of Mathematics.

[5] Choga, C.R. and Yu, S. (2026) 'A transformer-based multi-omics clinical decision support system for genomic risk stratification in lung adenocarcinoma', manuscript draft.

[6] CoFormerSurv Consortium (2026) 'CoFormerSurv: collaborative transformer for multi-omics survival analysis', PLOS Computational Biology.

[7] Collins, G.S. et al. (2024) 'TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods', BMJ, 385, e078378. doi: 10.1136/bmj-2023-078378.

[8] Cox, D.R. (1972) 'Regression models and life-tables', Journal of the Royal Statistical Society: Series B, 34(2), pp. 187-220.

[9] FDA (2026) Clinical Decision Support Software: Guidance for Industry and Food and Drug Administration Staff. Silver Spring, MD: U.S. Food and Drug Administration.

[10] FDA, Health Canada and MHRA (2025) Good Machine Learning Practice for Medical Device Development: Guiding Principles. Silver Spring, MD: U.S. Food and Drug Administration.

[11] Gao, J. et al. (2013) 'Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal', Science Signaling, 6(269), pl1. doi: 10.1126/scisignal.2004088.

[12] Hao, Y. et al. (2020) 'Integrated multi-omics analysis for cancer prognosis', Nature Communications, 11, pp. 1-12.

[13] Heagerty, P.J. and Zheng, Y. (2005) 'Survival model predictive accuracy and ROC curves', Biometrics, 61(1), pp. 92-105.

[14] Katzman, J.L. et al. (2018) 'DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network', BMC Medical Research Methodology, 18, 24. doi: 10.1186/s12874-018-0482-1.

[15] Lundberg, S.M. and Lee, S.-I. (2017) 'A unified approach to interpreting model predictions', Advances in Neural Information Processing Systems, 30.

[16] Mo, Q. et al. (2013) 'iClusterPlus: integrative clustering of multiple genomic data types', Nature Methods, 10(7), pp. 679-684. doi: 10.1038/nmeth.2651.

[17] Nhova, M., Mupa, M.N., Tsveta, L., Chingezi, L. and others (2026) 'Enterprise risk management in health coverage operations: Integrating premium billing controls-first analytics framework for eligibility-to-invoice integrity', World Journal of Advanced Research and Reviews.

[18] Okebu, D.T.C. and Mupa, M.N. (2026) 'Predictive training analytics for competency-based pilot instruction: Early detection of checkride risk, procedural noncompliance and safety-relevant learning gaps', World Journal of Advanced Research and Reviews.

[19] Pawlik, P. et al. (2025) 'Clone copy number diversity is linked to survival in lung cancer', Nature, 646(8083), pp. 190-197. doi: 10.1038/s41586-025-09398-w.

[20] Rappoport, N. and Shamir, R. (2019) 'Multi-omics integration: a comparative study', Genome Biology, 20, pp. 1-14.

[21] Sundararajan, M., Taly, A. and Yan, Q. (2017) 'Axiomatic attribution for deep networks', Proceedings of the 34th International Conference on Machine Learning, pp. 3319-3328.

[22] TCGA-LUAD Kaggle mirror (n.d.) TCGA-LUAD Dataset. Kaggle. Available at: https://www.kaggle.com/datasets/nahin333/tcgaluad (Accessed: 30 July 2026).

[23] Uno, H. et al. (2007) 'Evaluating prediction rules for survival endpoints', Statistics in Medicine, 26(22), pp. 3886-3904.

[24] Vaswani, A. et al. (2017) 'Attention is all you need', Advances in Neural Information Processing Systems, 30.

[25] Vickers, A.J. and Elkin, E.B. (2006) 'Decision curve analysis: a novel method for evaluating prediction models', Medical Decision Making, 26(6), pp. 565-574.

[26] Zehir, A. et al. (2017) 'Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients', Nature Medicine, 23, pp. 703-713. doi: 10.1038/nm.4333.

[27] Zhang, J. et al. (2021) 'Deep learning-based multi-omics integration for survival prediction in cancer', Nature Machine Intelligence, 3(12), pp. 1039-1048. doi: 10.1038/s42256-021-00427-8.