Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

PAST-Fusion: a leakage-controlled pathway-level multi-omics survival modeling framework in esophageal squamous cell carcinoma

Zhongyu Liu
The 989th Hospital of the Joint Logistics Support Force
Xue Wen
The 989th Hospital of the Joint Logistics Support Force
Hongwen Song
The 989th Hospital of the Joint Logistics Support Force
Yanjun Gao
The 989th Hospital of the Joint Logistics Support Force
Luojun Hui
ljhzzf241@163.com
The 989th Hospital of the Joint Logistics Support Force

Research Article

Keywords: esophageal squamous cell carcinoma, multi-omics, pathway-level modeling, survival analysis, nested cross-validation, external transportability

Posted Date: June 22nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9851307/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Multi-omics survival models for esophageal squamous cell carcinoma (ESCC) are vulnerable to data leakage, incomplete external molecular layers, and poor cross-cohort transportability. We developed PAST-Fusion as a leakage-controlled pathway-level framework for exploratory multi-omics prognostic modeling in ESCC.

PAST-Fusion provides a reproducible framework for exploratory pathway-level multi-omics modeling in ESCC. RNA+CNV showed the most stable internal signal in TCGA, but RNA-only deployment failed externally; therefore, the findings support future matched multi-omics validation rather than clinical deployment or validated biomarker discovery.

Conclusions

RNA+CNV showed the most stable internal performance in TCGA, whereas RNA-only showed weak signal and promoter methylation did not consistently improve performance beyond RNA+CNV. In GSE53625, the RNA-only deployment head failed to transport, with a continuous-risk C-index of 0.463 and failure of the TCGA-derived risk cutpoint. Diagnostic analyses indicated severe pathway-score distribution shift and risk-score translation failure.

Results

A TCGA ESCC fusion core cohort of 92 patients with matched RNA expression, copy-number variation, promoter methylation, and overall survival data was projected into a 50-Hallmark-pathway feature space. Ridge Cox models were evaluated using repeated nested cross-validation. GSE53625 was used only for locked RNA-only external deployment without external feature selection, scaling-parameter estimation, hyperparameter tuning, or cutpoint optimization.

Methods

Introduction

Esophageal squamous cell carcinoma (ESCC) remains a clinically challenging malignancy because patients with apparently similar clinicopathological features may experience substantially different survival outcomes. Conventional clinicopathological staging is essential for treatment planning, but it does not fully capture the molecular heterogeneity of ESCC. Molecular profiling has therefore become an important approach for exploring prognostic heterogeneity and for developing risk models that may support future biological stratification studies. [1] [2]

Multi-omics data provide complementary views of tumor biology. RNA expression reflects transcriptional activity, copy-number variation (CNV) captures genomic dosage imbalance, and promoter methylation represents an epigenetic regulatory layer. Integrating these data may provide a broader representation of tumor state than a single molecular view. However, direct gene-level integration is difficult in small matched cohorts because the number of molecular features can greatly exceed the number of patients. Pathway-level modeling offers one practical solution by projecting molecular data into a lower-dimensional biological feature space, improving interpretability and reducing the instability of purely gene-level feature selection. [3] [4]

A major methodological problem in public-cohort prognostic modeling is that apparent performance can be inflated by information leakage or by post-hoc optimization. This risk is particularly high when feature selection, scaling, hyperparameter tuning, or risk cutpoint selection is influenced by validation or external cohorts. External transcriptome cohorts also often lack matched CNV and methylation layers, making direct multi-omics deployment impossible. Under these conditions, an external dataset should not be treated as proof of model validation unless the deployment rule is locked before evaluation and no external optimization is performed. [5, 6] [7]

To address these issues, we developed PAST-Fusion as an exploratory, leakage-controlled, pathway-level framework for multi-omics survival modeling in ESCC. The framework was designed to compare RNA-only, RNA + CNV, and RNA + CNV+promoter methylation configurations within a matched TCGA ESCC fusion core cohort using repeated nested cross-validation, while reserving GSE53625 for locked RNA-only external deployment because that cohort contained transcriptome data but lacked matched CNV and methylation layers. The purpose was not to establish a clinically deployable prognostic signature, but to evaluate whether pathway-level multi-omics integration could improve internal modeling stability and to explicitly test the transportability limitations of an RNA-only deployment head in an external transcriptome-only cohort.

This study is worth reporting because it addresses a problem that is usually hidden in public-database prognostic modeling: an internally favorable model may collapse when the external cohort differs in platform, molecular-layer availability, and pathway-score distribution. PAST-Fusion was built to make that risk visible rather than to hide it. The same leakage-controlled workflow that identified an internal RNA + CNV signal also showed that the RNA-only deployment head failed in GSE53625 when no external tuning, cutpoint optimization, or risk-direction correction was allowed. The value of the work is therefore methodological and cautionary: it defines how exploratory multi-omics survival models should separate internal signal, external transportability, and future validation requirements. This study should not be read as evidence of external validation success, clinical utility, or validated biomarker discovery [5–7].

Methods

This study was designed as an exploratory, retrospective, pathway-level multi-omics survival modeling analysis in esophageal squamous cell carcinoma (ESCC). The workflow was developed to evaluate whether matched molecular layers could improve internal prognostic modeling while maintaining strict separation between model development, internal validation, and external transcriptome-only deployment.

The internal modeling cohort was a TCGA ESCC fusion core cohort consisting of 92 patients with valid overall survival information and matched molecular data, derived from TCGA/GDC resources and interpreted against the TCGA esophageal carcinoma molecular characterization framework [3, 8]. In the final recovered survival file, 31 events and 61 censored observations were retained. This TCGA fusion core cohort was used for model development, internal model comparison, and repeated nested cross-validation.

GSE53625 was used as an external transcriptome-only deployment cohort from the NCBI Gene Expression Omnibus [9, 10]. Because GSE53625 did not contain matched copy-number variation (CNV) or promoter methylation layers, it was not used to externally deploy the multi-omics RNA + CNV or RNA + CNV+promoter methylation configurations. It was used only for locked RNA-only external deployment assessment. GSE53625 contained 179 RNA-profiled ESCC samples after pathway-level processing.

The endpoint for survival modeling was overall survival. External deployment was treated as a transportability assessment, not as a model-development step. GSE53625 was not used for feature selection, model selection, hyperparameter tuning, scaling-parameter estimation, or risk-cutpoint optimization.

All molecular views were projected into a common Hallmark pathway feature space containing 50 pathway-level features based on the MSigDB Hallmark gene-set collection [4]. This representation was used to reduce dimensionality and to allow RNA expression, CNV, and promoter methylation to be compared within the same biological coordinate system.

At the rule level, RNA pathway activity was represented as a sample-level rank-based summary of genes belonging to each Hallmark gene set. CNV burden was summarized at the pathway level as the aggregate absolute copy-number burden of genes mapped to the same Hallmark gene set. Promoter methylation was summarized as a pathway-level promoter methylation state across mapped genes. These pathway-level summaries were used as modeling features rather than as gene-level biomarker claims.

For the TCGA fusion core cohort, three pathway-level feature matrices were used: RNA pathway activity, CNV absolute burden, and promoter methylation pathway state. The RNA matrix represented rank-based Hallmark pathway activity. The CNV matrix represented absolute pathway-level copy-number burden. The promoter methylation matrix represented pathway-level promoter methylation state. These matrices were aligned to the same TCGA fusion core sample set before modeling.

For GSE53625, only RNA pathway activity was available and processed into the corresponding Hallmark pathway-level feature space. No GSE53625 CNV or methylation feature matrices were available or imputed. Therefore, GSE53625 was restricted to RNA-only external deployment.

The primary pathway matrices used for Methods traceability were the frozen C01 v1 matrices identified in the repaired G02 source map. These frozen files were treated as the input feature layer for downstream modeling documentation. Alternative RNA pathway scoring matrices were treated as sensitivity or diagnostic inputs and were not used to replace the locked primary feature construction.

Survival models were implemented as ridge-regularized Cox proportional hazards models [11–13]. The candidate model configurations were defined before final performance comparison and included an RNA-only model, an RNA + CNV model, and an RNA + CNV+promoter methylation model. A baseline or null reference model was retained for comparison where specified in the modeling configuration.

Internal model evaluation in the TCGA fusion core cohort used repeated nested cross-validation, consistent with prediction-model reporting and validation principles [5–7]. The outer resampling loop was used to estimate internal predictive performance, whereas the inner loop was used for hyperparameter tuning. Hyperparameter selection was therefore restricted to the training portion of each outer split. Outer validation folds were not used for tuning.

The exact resampling, tuning, and model-configuration settings were defined before model execution in the project configuration files, including modeling_config.yaml and model_spec_baseline.yaml when present; these files were retained as the reproducibility source for the nested cross-validation and ridge-tuning specification. Hyperparameter tuning was kept inside the inner cross-validation loop, whereas the outer loop was reserved for internal performance estimation. No fold, repeat, or tuning choice was derived from GSE53625.

Preprocessing steps that could influence feature scaling or model fitting were performed within the appropriate training partitions and then applied to the corresponding validation partitions. This rule was used to avoid information leakage from validation samples into model fitting, feature scaling, or hyperparameter selection.

The main performance metric was the concordance index for continuous risk scores [14]. Model comparisons were interpreted as internal performance comparisons within TCGA only. The RNA + CNV configuration was not described as externally validated because no external matched multi-omics cohort was available for testing this configuration.

External deployment was restricted to the RNA-only model because GSE53625 contained transcriptome data but lacked matched CNV and promoter methylation data. The external analysis was therefore not a test of the RNA + CNV or RNA + CNV+promoter methylation models.

GSE53625 was excluded from all model-development operations. It was not used for feature selection, scaling-parameter estimation, hyperparameter tuning, model selection, or cutpoint optimization. The TCGA-derived modeling pipeline and training-defined risk rule were applied to the GSE53625 RNA pathway matrix as a locked external deployment procedure.

The external deployment was evaluated using continuous-risk prediction and the TCGA-derived training cutpoint where applicable. No GSE53625-specific cutpoint optimization was performed. If the training-defined cutpoint failed to stratify the external cohort, that failure was retained as part of the external deployment result.

External deployment diagnostics were used to interpret transportability failure. These diagnostics included pathway feature distribution shift between TCGA and GSE53625, coefficient-weighted feature shift, risk-score distribution shift, and risk-direction checks. Any risk-direction reversal analysis was treated as diagnostic only and was not used as a post-hoc correction or rescue of external validation.

Sensitivity analyses were performed to examine whether the observed internal model behavior and external deployment limitation depended on the primary RNA pathway scoring strategy. Alternative RNA pathway scoring branches included ssGSEA-based pathway activity and sample-rank-centered pathway activity [15]. A within-cohort gene z-score branch was also evaluated as a diagnostic branch.

These sensitivity analyses were separated from the primary analysis. Alternative scoring was not used to choose the primary model based on external performance. The within-cohort gene z-score branch used cohort-level distribution information and was therefore considered diagnostic rather than a locked external deployment transformation.

Sensitivity analyses were interpreted in relation to the primary leakage-controlled framework. They were used to assess the stability of the internal RNA + CNV finding and the persistence of cross-cohort distribution shift, not to convert the failed external deployment into external validation success.

All scripts were organized under the PAST_Fusion project directory and executed using R [16]. The G05 final-lock script was run under R version 4.3.0 with data.table version 1.17.0 [17]. The original model-training environment was traced through the C02/C03 modeling scripts, configuration files, QC outputs, and execution logs rather than recalculated during manuscript writing. The reproducibility package therefore includes the source maps, frozen pathway matrices, modeling configuration files, leakage-control plan, model-comparison plan, and external deployment diagnostic QC files needed to trace the Methods description.

Reproducibility was maintained through staged inputs, frozen pathway matrices, configuration files, and QC outputs. The repaired G02 source map identified the frozen pathway matrices, modeling configuration files, leakage-control plan, model-comparison plan, and external deployment diagnostic QC files used to support the Methods section.

The main leakage-control rules were as follows: TCGA and GSE53625 roles were defined before external evaluation; model tuning was performed only inside training partitions; GSE53625 was not used for feature selection, scaling-parameter estimation, hyperparameter tuning, or cutpoint optimization; and risk-direction reversal was treated as diagnostic only. These rules were applied to keep internal model development separate from external deployment assessment.

Results

Cohort freeze and pathway-level feature space

After cohort-level screening and sample harmonization, the TCGA ESCC fusion core cohort contained 92 patients with matched RNA expression, CNV, promoter methylation, and valid overall survival information. This cohort was used for model development and internal validation only. The external deployment cohort consisted of 179 GSE53625 ESCC samples with transcriptome and survival data, but without matched copy-number or methylation layers. Therefore, external deployment was restricted a priori to the transcriptome-only model head (Table 1; Figure 1).

Table 1. Cohort composition and pathway-level feature space

cohort_or_layer	role	sample_count	available_data	feature_space	analysis_use
TCGA fusion core	training and internal validation	92	RNA, CNV, promoter methylation, survival	50 MSigDB Hallmark pathways	Repeated nested-CV internal model development
GSE53625 deployment cohort	external transcriptome-only deployment diagnostic	179	RNA and survival	50 MSigDB Hallmark RNA pathway scores	M1 RNA-only external continuous-risk evaluation
RNA pathway layer	primary and sensitivity feature layer	TCGA 92 / GSE53625 179	Gene expression transformed to Hallmark pathway scores	S0 rank-mean, S1 ssGSEA, S2 rank-centered, S3 diagnostic z-score	Internal model comparison and external deployment diagnostics
CNV pathway layer	TCGA-only multi-omics feature layer	92	CNV absolute pathway burden	50 Hallmark pathways	Internal M2/M3 models only
Methylation pathway layer	TCGA-only multi-omics feature layer	92	Promoter methylation pathway state	50 Hallmark pathways	Internal M3 model only

Footnote: TCGA fusion core was used for internal model development and repeated nested cross-validation. GSE53625 was transcriptome-only and was used only for RNA-only external deployment evaluation. CNV and promoter methylation layers were unavailable in GSE53625; therefore, RNA+CNV and RNA+CNV+methylation models were not externally deployed.

To reduce feature heterogeneity across omics layers, all molecular measurements were projected into a shared Hallmark pathway-level feature space. RNA expression was represented as pathway activity, copy-number alteration as pathway-level CNV burden, and promoter methylation as pathway methylation state. This resulted in three aligned 50-pathway feature matrices for the TCGA fusion core cohort and a transcriptome-only 50-pathway matrix for GSE53625. The matched TCGA feature matrices had identical sample and pathway orders, allowing direct comparison of RNA-only, RNA+CNV, and RNA+CNV+methylation model configurations (Table 1; Figure 1).

Leakage-controlled internal validation framework

All survival models were evaluated under a repeated nested cross-validation design. The outer loop was used only for internal performance estimation, while the inner loop was used for hyperparameter tuning. Feature scaling parameters were estimated within the relevant training partition and then applied unchanged to the corresponding validation partition. No global pre-scaling, external cohort feature selection, external hyperparameter tuning, or GSE53625 cutpoint optimization was allowed (Figure 1).

The baseline model set was intentionally kept small. Four configurations were evaluated in the primary analysis: a null baseline model, an RNA-only ridge Cox model, an RNA+CNV ridge Cox model, and an RNA+CNV+promoter methylation ridge Cox model. The RNA-only model was also the only model eligible for GSE53625 deployment, because the external cohort did not contain CNV or methylation data (Table 1; Figure 1).

Internal nested cross-validation showed the most stable signal for RNA+CNV

In the primary internal analysis using the S0 rank-based mean RNA pathway activity score, the null baseline model produced the expected C-index of 0.500, confirming that the evaluation pipeline behaved as expected. The RNA-only ridge

Cox model showed only weak internal prognostic signal, with a mean repeated out-of-fold C-index of 0.518 (Table 2; Figure 2).

Table 2. Internal and external performance of pathway-level ridge Cox models

analysis_block	rna_scoring	model	mean_cindex	median_cindex	role	interpretation
Primary internal nested CV	S0 rank-mean	M1 RNA-only	0.518	0.508	Internal baseline	Weak RNA-only internal signal
Primary internal nested CV	S0 rank-mean	M2 RNA+CNV	0.570	0.574	Primary internal best model	Most stable internal improvement
Primary internal nested CV	S0 rank-mean	M3 RNA+CNV+METH	0.556	0.557	Three-modality exploratory model	Did not improve over RNA+CNV
External deployment	S0 rank-mean	M1 RNA-only	0.463	NA	Only externally deployable model	Failed external transportability
Internal sensitivity	S1 ssGSEA	M2 RNA+CNV	0.568	0.567	Sensitivity	RNA+CNV remained best under S1
Internal sensitivity	S2 rank-centered	M2 RNA+CNV	0.591	0.602	Sensitivity	Highest internal sensitivity result but not primary replacement
Diagnostic sensitivity	S3 gene-zscore	M3 RNA+CNV+METH	0.557	0.576	Diagnostic only	Cannot redefine primary model

Footnote: Internal C-index values are repeated out-of-fold estimates from the TCGA fusion core cohort. The GSE53625 value represents continuous-risk external deployment of the RNA-only model, not external validation success. The external deployment row has no median C-index because it represents a single external cohort evaluation.

Adding pathway-level CNV burden improved internal performance. The RNA+CNV model achieved the highest mean repeated out-of-fold C-index in the primary analysis, with a mean C-index of 0.570 and a median C-index of 0.574. In contrast, adding promoter methylation to RNA+CNV did not further improve performance. The RNA+CNV+promoter methylation model achieved a mean C-index of 0.556, which remained above the RNA-only model but below the RNA+CNV model (Table 2; Figure 2).

The paired repeated-CV comparisons supported the same pattern. Across repeated splits, RNA+CNV outperformed RNA-only in most repetitions, whereas RNA+CNV+promoter methylation did not consistently outperform RNA+CNV. These findings indicate that, within the TCGA fusion core cohort, adding pathway-level CNV burden yielded the most stable internal predictive gain beyond RNA pathway activity, whereas the methylation layer did not show a robust additional prognostic contribution under the current pathway-level construction (Figure 2).

RNA-only external deployment failed in GSE53625

Because GSE53625 contained transcriptome data only, external deployment was restricted to the RNA-only ridge Cox model. The final RNA-only deployment model was trained using TCGA-derived preprocessing and model parameters, and

was then applied to GSE53625 without external feature selection, external scaling-parameter estimation, external hyperparameter tuning, or external cutpoint optimization (Figure 1).

The apparent C-index of the TCGA-trained RNA-only model in the training cohort was 0.804. However, this apparent performance did not translate to the external transcriptome-only cohort. In GSE53625, the continuous-risk C-index was 0.463, indicating failed external transportability rather than successful external validation (Table 2; Figure 3).

The training-defined median risk cutpoint also failed to stratify the external cohort. When the TCGA-derived cutpoint was applied unchanged to GSE53625, no external samples were assigned to the high-risk group, and all 179 samples fell below the training-defined threshold. Therefore, threshold-based external Kaplan-Meier stratification was not interpretable as a valid external validation result (Table 3; Figure 3).

Table 3. Diagnostics of RNA-only external deployment failure in GSE53625

diagnostic_domain	metric	value	interpretation
External performance	GSE53625 continuous-risk C-index	0.463	Failed external transportability
Training optimism	TCGA apparent C-index	0.804	Apparent training performance is much higher than external performance
Cutpoint translation	GSE above / below training-defined cutpoint	0 / 179	Training cutpoint failed in GSE53625
Direction check	Direction reversal flag	TRUE	Diagnostic only; not a valid correction
Direction check	Reversed C-index diagnostic only	0.537	Cannot be used as post-hoc external validation
Distribution shift	Failure mechanism	TCGA-GSE pathway score distribution shift with risk-score translation failure	Main explanation for external deployment failure

Footnote: The direction-reversal result is diagnostic only and was not used as a model correction. The training-defined cutpoint was applied unchanged to GSE53625 and failed to stratify the external cohort. No GSE53625 cutpoint optimization was performed.

External failure was associated with pathway-score distribution shift and risk-score translation failure

To investigate why the RNA-only deployment head failed externally, we performed diagnostic analyses of pathway-score and risk-score distributions. The diagnostics showed that the GSE53625 pathway activity profiles were not aligned with the TCGA-derived feature space after applying the locked TCGA-trained transformation and model (Table 3; Figure 4).

All 50 Hallmark pathways showed marked distribution shift between TCGA and GSE53625. The median absolute standardized mean difference across pathways was 9.165, and all 50 pathways had an absolute standardized mean difference greater than 2. This degree of shift indicated that the external cohort did not occupy the same pathway-score distribution as the TCGA training cohort (Table 3; Figure 4).

The magnitude and direction of this pathway-score shift were consistent with the observed displacement of the GSE53625 risk-score distribution. The GSE53625 risk-score distribution was shifted far below the TCGA training risk-score distribution,

explaining why the TCGA-derived cutpoint assigned all external samples to the low-risk group. Directional diagnostics further showed that reversing the external risk-score direction yielded a C-index of 0.537. However, this result was treated strictly as a diagnostic signal and was not used as a post-hoc correction, because risk-direction reversal based on external performance would violate the locked external deployment rule (Table 3; Figure 3).

Together, these results indicate that the external deployment failure was not caused by an inability to generate risk scores, but by a failure of risk-score transportability across cohorts and platforms.

Alternative RNA pathway scoring did not overturn the internal conclusion

We next examined whether the internal findings depended on the original RNA pathway scoring method. Three pre-specified alternative RNA scoring approaches were evaluated: ssGSEA-based scoring, sample-rank-centered scoring, and within-cohort gene z-score pathway scoring. These analyses were conducted as sensitivity or diagnostic analyses and did not replace the primary S0 analysis (Figure 4).

Under ssGSEA-based scoring, the RNA+CNV model again achieved the best internal performance among the evaluated configurations, with a mean repeated out-of-fold C-index of 0.568. Under sample-rank-centered scoring, the RNA+CNV model reached the highest internal sensitivity result, with a mean C-index of 0.591. In both branches, the addition of promoter methylation did not improve performance beyond RNA+CNV (Table 2; Figure 2).

The within-cohort gene z-score branch showed a different model ranking, with the RNA+CNV+promoter methylation model producing the highest C-index within that diagnostic branch. However, this result was not used to redefine the primary model, because within-cohort z-score scoring uses cohort-level distribution information and does not represent a fully locked deployment transformation (Figure 4).

Importantly, the alternative scoring analyses did not materially resolve the TCGA-GSE pathway-score distribution shift. ssGSEA and sample-rank-centered scoring still showed severe cross-cohort distribution differences. Thus, the external deployment failure was not explained simply by the initial S0 rank-based pathway scoring method (Figure 4).

Integrated interpretation of modeling results

Across the primary analysis and sensitivity analyses, the most stable internal finding was that pathway-level RNA+CNV fusion outperformed RNA-only modeling within the TCGA fusion core cohort. Promoter methylation did not provide consistent additional predictive gain beyond RNA+CNV in the current sample size and feature construction (Table 2; Figure 2).

In contrast, the RNA-only deployment head did not transport successfully to GSE53625 under strict no-leakage evaluation. The failure was associated with severe cross-cohort pathway-score distribution shift and risk-score translation failure. Therefore, the current results support PAST-Fusion as a reproducible, leakage-controlled, pathway-level exploratory framework, but do not support claims of external validation, clinical utility, or validated biomarker discovery (Table 2; Table 3).

Discussion

This study developed PAST-Fusion as a leakage-controlled, pathway-level framework for exploratory multi-omics survival modeling in ESCC. Its main value is not that it delivers a clinically deployable prognostic signature; it does not. Its value is that it exposes a common weakness of public-cohort prognostic studies: internal performance can look acceptable even

when the model cannot be transported to an external cohort with different platform characteristics and incomplete molecular layers. Under repeated nested cross-validation, pathway-level RNA + CNV fusion showed the most stable internal signal in the TCGA ESCC fusion core cohort. Under a locked external deployment rule, however, the RNA-only deployment head failed in GSE53625. This negative external result is central to the manuscript. It shows why internal performance, even when estimated with nested validation, should not be upgraded to claims of external validation or clinical utility without an independent matched multi-omics cohort [5–7].

The first major finding was that adding pathway-level CNV burden to RNA pathway activity improved internal model performance in the TCGA fusion core cohort. The RNA-only model showed weak internal prognostic signal, whereas the RNA + CNV model achieved the most stable performance across the primary repeated nested cross-validation analysis. This pattern was broadly consistent in sensitivity analyses using ssGSEA-based and sample-rank-centered RNA pathway scoring. These results suggest that, within the matched TCGA multi-omics setting, CNV burden provided a useful complementary pathway-level representation for internal prediction. This interpretation, however, should remain predictive rather than causal. The present analysis supports an internal modeling gain from adding CNV features; it does not establish that CNV alterations are direct biological drivers of prognosis, nor does it validate individual CNV-affected pathways as biomarkers.

The promoter methylation layer did not provide a consistent additional improvement beyond RNA + CNV. This should not be interpreted as evidence that methylation is biologically irrelevant in ESCC. Rather, within the current pathway-level construction and limited sample size, adding promoter methylation to RNA + CNV did not yield a stable predictive gain. Several explanations are possible. The matched fusion core cohort was small, which limited the ability to estimate incremental effects from a third omics layer. Promoter methylation effects may also be more gene-specific or context-dependent than the pathway-level summary used here can capture. In addition, part of the methylation signal may overlap with transcriptional and copy-number features already represented in the RNA and CNV layers. Therefore, the appropriate conclusion is limited: this pathway-level methylation representation did not improve the internal survival model beyond RNA + CNV in the present analysis.

A central feature of this study is that external deployment was treated as a strict transportability assessment rather than as an opportunity for post-hoc optimization. Because GSE53625 contains transcriptome data but lacks matched CNV and methylation layers, only the RNA-only model could be externally deployed. Under this locked evaluation, the RNA-only deployment head did not generalize to GSE53625. The external continuous-risk C-index was below 0.5, and the TCGA-derived training cutpoint failed to stratify the external cohort. This result should be interpreted as a negative transportability finding. It should not be reframed as successful external validation, and it should not be corrected by reversing the risk direction, optimizing a GSE-specific cutpoint, or selecting an alternative pathway scoring method based on external performance.

The diagnostic analyses suggest that the external deployment failure was related to cross-cohort pathway-score distribution shift and risk-score translation failure. The GSE53625 pathway activity profiles showed substantial distributional differences from the TCGA-derived pathway-score space. This shift was observed across all Hallmark pathways and was consistent with displacement of the external risk-score distribution. Thus, the model could generate numerical risk scores in GSE53625, but the scale and orientation of those scores did not preserve the TCGA-derived risk structure. The failure was therefore not simply an unfavorable Kaplan-Meier separation problem; it reflected a deeper limitation in transferring a TCGA-trained pathway-level risk model to a transcriptome-only microarray cohort.

The alternative RNA pathway scoring analyses reinforced this interpretation. If the failed deployment had been mainly caused by the original S0 rank-based pathway scoring method, alternative scoring strategies would be expected to reduce the TCGA-GSE distribution shift. This was not observed. ssGSEA-based and sample-rank-centered scoring did not materially resolve the cross-cohort shift, although sample-rank-centered scoring slightly improved internal TCGA performance. The within-cohort gene z-score branch reduced mean-level differences by construction, but it used cohort-level distribution

information and was therefore treated as diagnostic only. It cannot be considered a locked external deployment transformation. These findings indicate that external transportability cannot be assumed to improve simply by changing the pathway scoring method.

The methodological contribution of PAST-Fusion lies in its strict separation of model development, sensitivity analysis, and external deployment diagnostics. The fusion core cohort was frozen, the pathway feature space was fixed, the repeated nested cross-validation design was specified, and GSE53625 was excluded from model tuning, scaling-parameter estimation, feature selection, and cutpoint optimization. This design reduced common sources of information leakage and made the negative external result more informative. The failed deployment could not be hidden by retrospectively changing the cutpoint, reversing the risk direction, or choosing the best-performing external scoring branch. In this sense, the study provides a reproducible framework for evaluating both internal multi-omics signal and cross-cohort transportability limits.

Another strength of the framework is the use of pathway-level representation across multiple molecular layers. Pathway-level modeling reduces dimensionality and allows RNA, CNV, and methylation data to be represented in a common biological coordinate system. This is useful when sample size is limited and direct gene-level multi-omics fusion would be unstable. It also allows clearer comparison between RNA-only, RNA + CNV, and RNA + CNV+methylation configurations. However, pathway-level summarization has limitations. It may obscure gene-level heterogeneity, collapse opposing effects within the same pathway, and reduce sensitivity to localized molecular events. Therefore, the pathway-level results should be interpreted as modeling constructs and hypothesis-generating signals, not as evidence that specific pathways are validated biomarkers or therapeutic targets.

This study has several important limitations. First, the TCGA fusion core cohort included only 92 matched ESCC cases, which is small for multi-omics survival modeling even with pathway-level dimensionality reduction and ridge regularization. Repeated nested cross-validation provides a more conservative internal estimate than apparent training performance, but it cannot compensate for limited cohort size and event numbers. Second, external deployment was necessarily limited to the RNA-only model because GSE53625 lacked matched CNV and methylation data. As a result, the strongest internal configuration, RNA + CNV, could not be externally tested in a matched multi-omics cohort. Third, GSE53625 differs from TCGA in platform and pathway-score distribution, so the external result reflects a combined challenge of cohort transfer, platform transfer, and transcriptome-only deployment.

The study also lacks independent matched multi-omics validation, prospective validation, and wet-lab confirmation. Without an external ESCC cohort containing RNA, CNV, methylation, and survival data processed under a predefined pipeline, it is not possible to determine whether the internal RNA + CNV advantage would remain stable outside TCGA. The present results do not validate candidate pathways as biomarkers or functional drivers, and they do not support direct clinical application. Any future clinical interpretation would require independent cohort validation, calibration assessment, and ideally prospective or multicenter evaluation.

Future work should proceed along a stricter validation path rather than attempting to rescue the current GSE53625 result. The most important next step is to assemble or access independent ESCC cohorts with matched RNA, CNV, methylation, and survival information. In such cohorts, pathway scoring, feature scaling, hyperparameter tuning, and risk-score calculation should be predefined before testing. RNA + CNV and RNA + CNV+methylation models should then be evaluated as locked models in an external matched multi-omics setting. Continuous-risk evaluation should remain the primary external endpoint, and threshold-based stratification should rely on training-defined thresholds unless a separate validation design explicitly permits recalibration.

Future studies should also incorporate formal domain-shift diagnostics before claiming transportability. The present work shows that a model may appear reasonable internally but fail externally because the target cohort occupies a different pathway-score distribution. External validation should therefore not be reduced to a single C-index or Kaplan-Meier p-value. It should include feature-distribution diagnostics, risk-score calibration, platform-harmonization assessment, and

predefined rules for handling scale mismatch. If candidate pathways are prioritized from the model, they should be tested in independent molecular or functional experiments before being described as biomarkers or targets.

Overall, PAST-Fusion should be viewed as an exploratory, leakage-controlled framework for pathway-level multi-omics survival modeling in ESCC. The internal results suggest that RNA + CNV fusion may provide a more stable prognostic signal than RNA-only modeling within matched TCGA data, while promoter methylation did not add consistent predictive value in the present setting. At the same time, the failed RNA-only deployment in GSE53625 highlights a major transportability barrier for public-cohort prognostic modeling. The current findings therefore support methodological feasibility and define a clear future validation path, but they do not support claims of external validation success, clinical utility, or validated biomarker discovery.

Conclusion

PAST-Fusion was developed as an exploratory, leakage-controlled framework for pathway-level multi-omics survival modeling in ESCC. In the TCGA fusion core cohort, RNA + CNV showed the most stable internal prognostic signal, whereas promoter methylation did not provide consistent incremental gain under the current pathway-level construction. However, the RNA-only deployment head failed to transport to GSE53625, indicating a major cross-cohort transportability limitation associated with pathway-score distribution shift and risk-score translation failure. These findings support the feasibility of a structured pathway-level multi-omics modeling workflow and provide a basis for future matched multi-omics validation, but they do not support clinical deployment, external validation claims, or validated biomarker discovery.

Abbreviations

CNV: copy-number variation

C-index: concordance index

ESCC: esophageal squamous cell carcinoma

GEO: Gene Expression Omnibus

GSE: Gene Expression Series

MSigDB: Molecular Signatures Database

OS: overall survival

PAST-Fusion: Pathway-level multi-omics survival modeling framework used in this study

PROBAST: Prediction model Risk Of Bias ASsessment Tool

ssGSEA: single-sample gene set enrichment analysis

TCGA: The Cancer Genome Atlas

TRIPOD: Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis

Declarations

Ethics approval and consent to participate

This study was based on secondary analysis of publicly available, de-identified TCGA and GEO datasets. No new human participants were recruited, and no experiments or interventions were performed by the authors. According to the policies of The Cancer Genome Atlas and the Gene Expression Omnibus, use of these de-identified data does not require additional ethics approval or informed consent. The study was conducted in accordance with the Declaration of Helsinki.

Consent for publication

Not applicable. This manuscript does not contain identifiable individual-level information.

Availability of data and materials

The datasets analyzed in this study are publicly available. TCGA esophageal carcinoma molecular and clinical data were obtained from The Cancer Genome Atlas / Genomic Data Commons (TCGA/GDC), project identifier TCGA-ESCA, available through the GDC Data Portal: https://portal.gdc.cancer.gov/projects/TCGA-ESCA. The external transcriptome cohort was obtained from the Gene Expression Omnibus (GEO), dataset title “LncRNA profile study reveals a three-lncRNA signature associated with the survival of esophageal squamous cell carcinoma patients”, accession number GSE53625, available at: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE53625 and https://identifiers.org/geo:GSE53625. Hallmark pathway gene sets used for pathway-level feature construction were obtained from the Molecular Signatures Database (MSigDB) Hallmark collection, available at: https://www.gsea-msigdb.org/gsea/msigdb/human/collections.jsp#H.

The processed cohort summaries, pathway-level feature-construction summaries, nested cross-validation/modeling summaries, internal and external performance outputs, external deployment diagnostic summaries, and RNA pathway scoring sensitivity summaries generated during this study are provided in the Supplementary Information as Supplementary Data S1–S6. Additional intermediate analysis files and scripts are available from the corresponding author upon reasonable request, subject to institutional data-sharing and repository-preparation requirements.

Competing interests

The authors declare that they have no competing interests.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Authors' contributions

ZYL conceived and designed the study, performed the computational analysis, developed the analysis workflow, and drafted the manuscript. XW contributed to study design, resources, and manuscript revision. JHL contributed to data organization, methodology review, and manuscript revision. YJG contributed to clinical interpretation, supervision, and manuscript revision. HWS contributed to supervision, data curation, validation of the analysis workflow, and manuscript revision. All authors read and approved the final manuscript.

Acknowledgements

The authors thank the TCGA and GEO research communities for making the data publicly available.

References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74(3):229–263. doi:10.3322/caac.21834. https://doi.org/10.3322/caac.21834
2. Song Y, Li L, Ou Y, Gao Z, Li E, Li X et al. Identification of genomic alterations in oesophageal squamous cell cancer. Nature. 2014;509(7498):91–95. 10.1038/nature13176. URL: https://www.nature.com/articles/nature13176
3. Cancer Genome Atlas Research Network. Integrated genomic characterization of oesophageal carcinoma. Nature. 2017;541(7636):169–175. 10.1038/nature20805. URL: https://www.nature.com/articles/nature20805
4. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database Hallmark Gene Set Collection. Cell Syst. 2015;1(6):417–25. 10.1016/j.cels.2015.12. .004. URL.
5. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS et al. PROBAST: A tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med. 2019;170(1):51–58. doi:10.7326/M18-1376. https://doi.org/10.7326/M18-1376
6. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD statement. Ann Intern Med. 2015;162(1):55–63. 10.7326/M14-0697. URL: https://doi.org/10.7326/M14-0697.
7. Steyerberg EW, Harrell FE Jr. Prediction models need appropriate internal, internal-external, and external validation. J Clin Epidemiol. 2016;69:245–7. 10.1016/j.jclinepi.2015.04.005. URL.
8. Grossman RL, Heath AP, Ferretti V, Varmus HE, Lowy DR, Kibbe WA et al. Toward a shared vision for cancer genomic data. N Engl J Med. 2016;375(12):1109–1112. 10.1056/NEJMp1607591. URL: https://doi.org/10.1056/NEJMp1607591.
9. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M et al. NCBI GEO: archive for functional genomics data sets-update. Nucleic Acids Res. 2013;41(Database issue):D991-D995. 10.1093/nar/gks1193. URL: https://doi.org/10.1093/nar/gks1193.
10. National Center for Biotechnology Information. GEO Series GSE53625. LncRNA profile study reveals a three-lncRNA signature associated with the survival of esophageal squamous cell carcinoma patients. GEO Accession GSE53625. Accessed 2026 May 28. URL: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE53625
11. Cox DR. Regression models and life-tables., Stat Soc Series JR. B Stat Methodol. 1972;34(2):187–220. 10.1111/j.2517-6161.1972.tb00899.x. URL: https://doi.org/10.1111/j.2517-6161.1972.tb00899.x.
12. Simon N, Friedman J, Hastie T, Tibshirani R. Regularization paths for Cox's proportional hazards model via coordinate descent. J Stat Softw. 2011;39(5):1–13. 10.18637/jss.v039.i05. URL: https://www.jstatsoft.org/article/view/v039i05
13. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw. 2010;33(1):1–22. 10.18637/jss.v033.i01. URL: https://www.jstatsoft.org/article/view/v033i01
14. Harrell FE Jr, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. JAMA. 1982;247(18):2543–2546. URL: https://pubmed.ncbi.nlm.nih.gov/7069920/
15. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics. 2013;14:7. 10.1186/1471-2105-14-7. URL: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-14-7
16. R Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2023. https://www.R-project.org/.
17. Dowle M, Srinivasan A. data.table: Extension of data.frame. R package version 1.17.0. Accessed. 2026 May 28. URL: https://CRAN.R-project.org/package=data.table

Figures

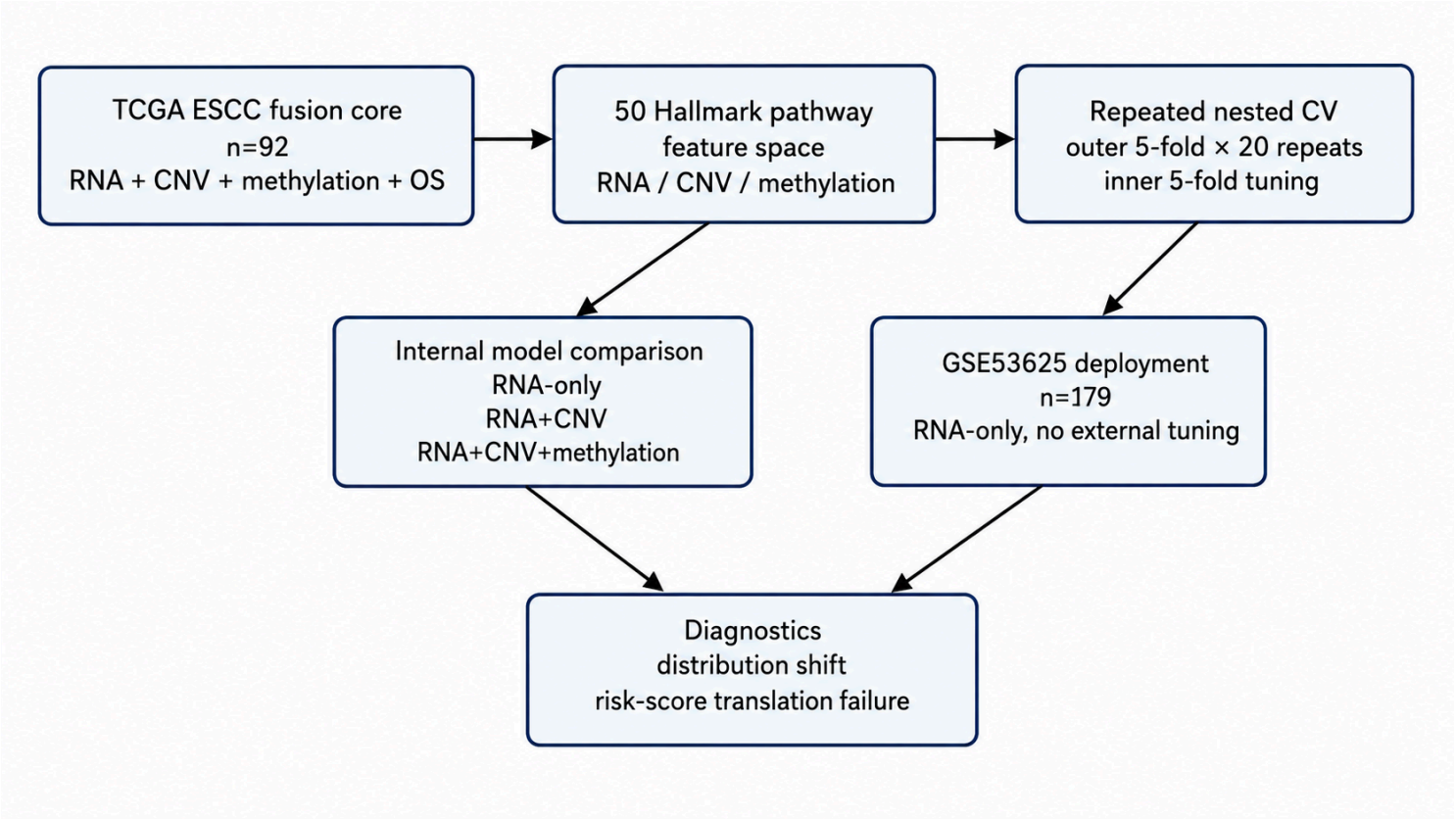

Figure 1

PAST-Fusion workflow and leakage-controlled analysis design. The TCGA fusion core cohort included 92 ESCC patients with matched RNA, CNV, promoter methylation, and survival information. Molecular features were projected to a common 50-Hallmark-pathway space. Internal model development used repeated nested cross-validation with fold-specific scaling. GSE53625 contained transcriptome data only and was restricted to RNA-only external deployment. No external feature selection, external hyperparameter tuning, external scaling-parameter estimation, or GSE53625 cutpoint optimization was performed.

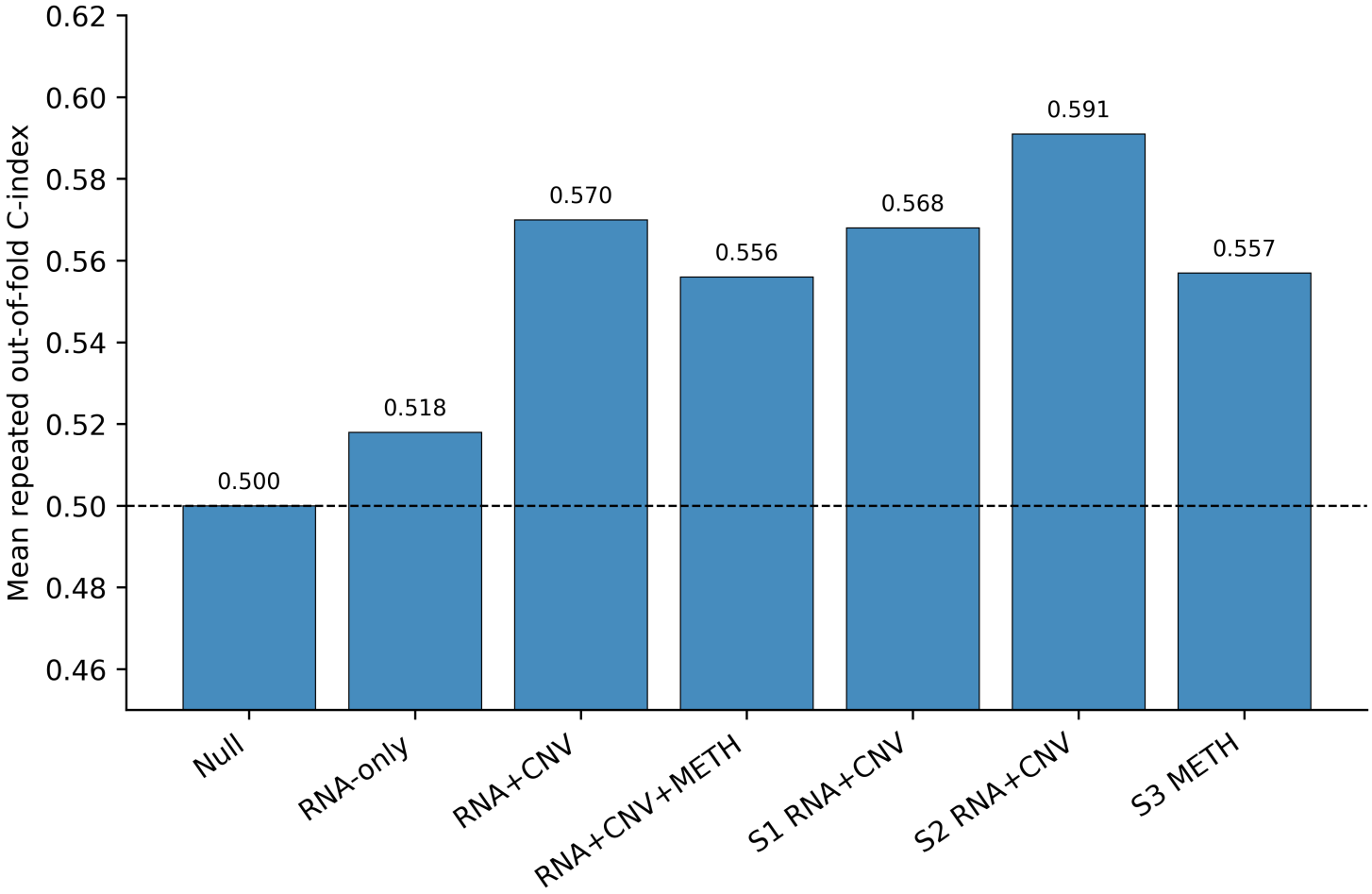

Figure 2

Internal repeated nested-CV model comparison. Under the primary rank-based RNA pathway scoring, the RNA+CNV ridge Cox model showed the most stable internal performance. The RNA+CNV+methylation model did not consistently outperform RNA+CNV. Alternative RNA scoring sensitivity analyses using ssGSEA and sample-rank-centered scoring did not overturn the internal RNA+CNV finding.

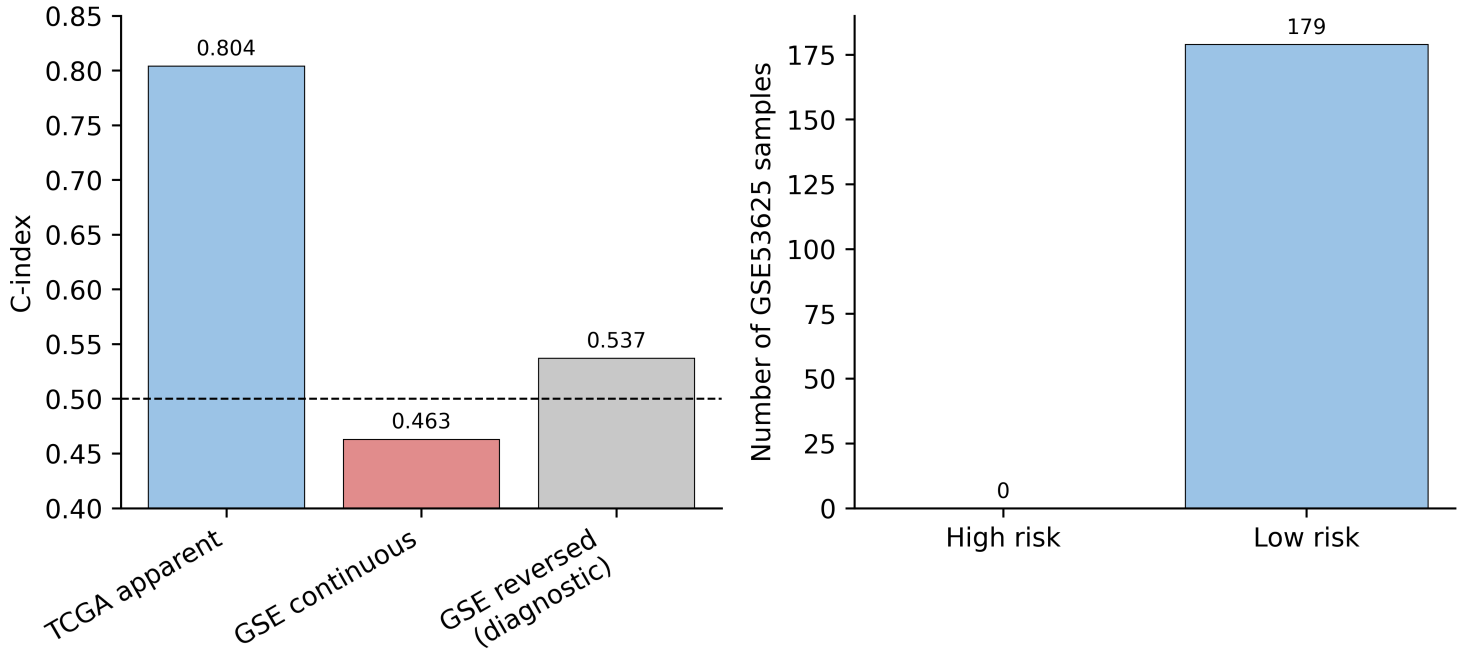

Figure 3

RNA-only external deployment failure in GSE53625. The RNA-only deployment head was trained in the TCGA fusion core cohort and applied unchanged to GSE53625. The external continuous-risk C-index was 0.463, indicating failed external transportability. The TCGA training-defined cutpoint assigned 0 GSE53625 samples to the high-risk group and 179 samples to the low-risk group. Risk-direction reversal yielded a diagnostic C-index of 0.537 but was not used as a post-hoc model correction.

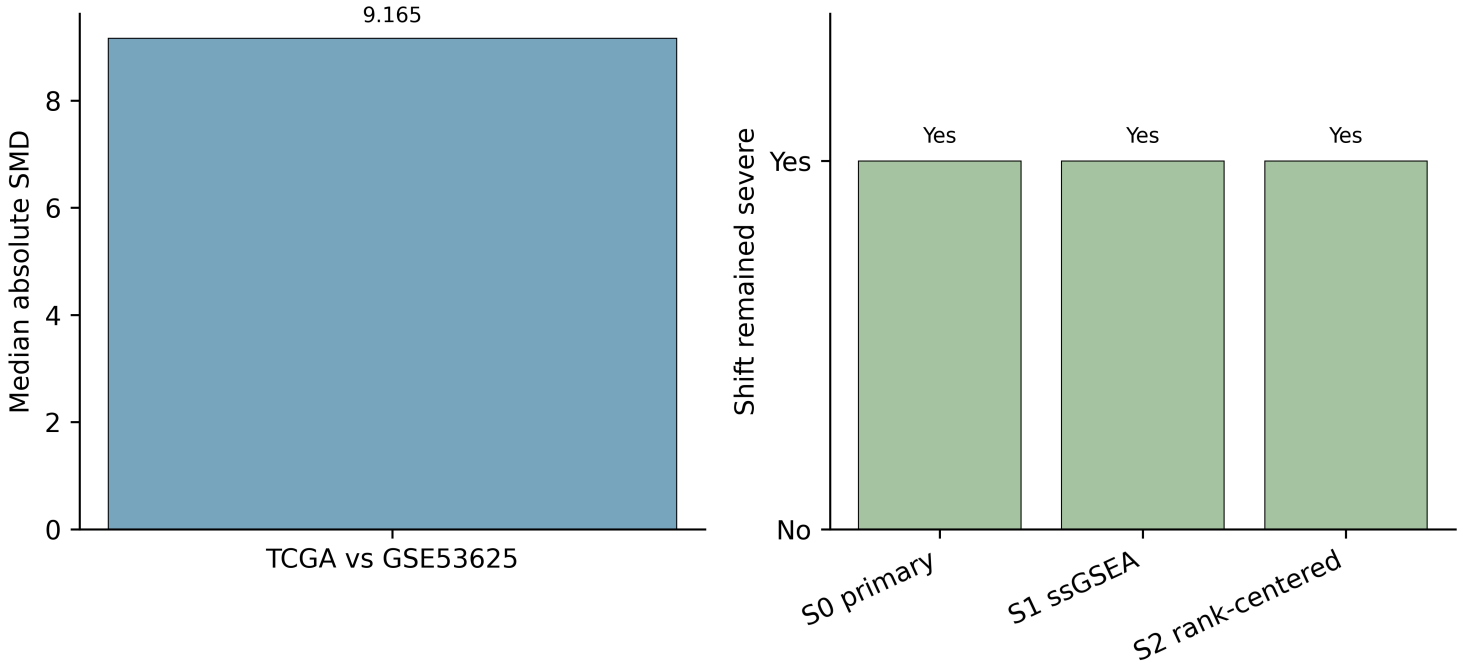

Figure 4

Alternative RNA pathway scoring and TCGA-GSE distribution shift. Alternative scoring methods did not materially resolve the TCGA-GSE pathway-score distribution shift or overturn the primary internal conclusion that RNA+CNV was the most

stable configuration.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterialsCaptionandIndex.docx
- SupplementaryMaterialsDiscoverOncology.xlsx