GLOBAL-TO-INSTANCE STRUCTURAL ALIGNMENT FOR PATHWAY-BASED MULTI-VIEW LEARNING IN MULTIMODAL CANCER SURVIVAL PREDICTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Multimodal cancer survival prediction aims to leverage complementary information from histopathological images and gene expression to estimate patient-specific prognostic risks. Recent approaches organize genomic information into biological pathways, forming a multi-view learning framework where each pathway serves as an individual view jointly modeled with the shared histopathological modality. However, these methods often treat pathways as independent sources, overlooking complex inter-view relationships among biological pathways. We investigate pathway structures from a population perspective and identify heterogeneous cross-pathway correlations and pathway-survival associations across cancer cohorts. Motivated by these observations, we propose StructAlign, a global-to-instance alignment framework that incorporates population-level pathway structures into patient-specific multi-view representation learning. StructAlign introduces cross-pathway and pathway-survival alignment objectives to preserve inter-pathway dependencies and guide pathway aggregation toward survival-associated patterns. Extensive experiments across multiple cancer cohorts demonstrate that StructAlign improves survival prediction and interpretability through consistent pathway structure preservation, survival-aware view selection, and biologically meaningful attribution.

1 INTRODUCTION

Histopathological whole slide images (WSIs) provide rich morphological and spatial information of tumor tissues (Song et al., 2023; Lu et al., 2021), while gene expression profiles characterize molecular alterations associated with tumor progression (El Nahhas et al., 2024; Zeng et al., 2023). By jointly modeling these complementary and heterogeneous modalities, **multimodal cancer survival prediction** has emerged as a promising direction for precision oncology, aiming to estimate patient-specific prognostic risks (Győrffy, 2021; Afonso et al., 2024). Recent studies further incorporate biological knowledge by grouping genes into functional categories, such as pathways (Jaume et al., 2024; Yan et al., 2025). This naturally leads to a **multi-view learning** setting, where each pathway serves as an individual biological view that interacts with shared histopathological features under its specific biological context, enabling the characterization of tumor morphology from distinct functional perspectives.

Although pathway-based methods formulate the problem as multi-view learning, they often treat pathway views as isolated sources (Xu & Chen, 2023; Zhang et al., 2024), focusing on view-specific representations and simple fusion while overlooking *intrinsic inter-view relationships* essential for effective multi-view learning (Xu et al., 2024). In reality, biological pathways form interconnected systems with complex dependencies and heterogeneous clinical implications (Hanahan & Weinberg, 2011; Sanchez-Vega et al., 2018). Therefore, preserving *globally consistent pathway structures* is crucial for biologically coherent representation learning and reliable pathway-aware risk prediction.

To investigate pathway-level structures, we analyze the pathway relationships from a population perspective. Results reveal that pathways are not independent biological signals; instead, they exhibit structured **cross-pathway correlation** patterns, where groups of pathways share similar activity profiles across patients. Furthermore, the **pathway-survival associations** vary substantially across

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

cancer types, with different pathways exhibiting risk-promoting or protective relationships depending on cancer context. These findings demonstrate that pathway views contain complementary inter-view structures beyond view-specific information, whose preservation is crucial for effective multi-view representation learning by enabling models to jointly exploit pathway-specific characteristics and underlying relationships across pathways.

Motivated by these observations, we propose **StructAlign**, a global-to-instance structural alignment framework for pathway-based multi-view learning. Different from existing approaches that independently model pathway views, StructAlign incorporates population-level pathway structures as biological guidance for patient-specific representation learning. To preserve the structural relationships among pathway views, we introduce a **Cross-Pathway Alignment** objective to align pathway-specific multimodal representations, where each pathway view characterizes histopathological features from its own biological perspective while capturing shared structural information across views. Building upon the survival-associated patterns, we further develop a **Pathway-Survival Alignment** objective to align patient-specific pathway aggregation with population-level prognostic structures. Through these constraints, StructAlign bridges population-level knowledge and instance-level multimodal representation learning, enabling structure-aware multi-view learning with biologically consistent pathway representations.

Extensive experiments across six cancer cohorts from The Cancer Genome Atlas (TCGA) (Weinstein et al., 2013) demonstrate that StructAlign consistently outperforms state-of-the-art methods. Further structural consistency analysis, survival-associated pathway preference analysis, and attribution visualization validate that StructAlign effectively preserves meaningful pathway structures and improves pathway-level interpretability. Our main contributions are as follows:

1. **Unveiling Pathway-aware Structural Organizations.** We analyze heterogeneous correlations and prognostic associations among pathways, motivating a structural guidance for pathway-based multi-view learning.
2. **Global-to-instance structural alignment.** We propose StructAlign, a structural alignment framework that incorporates population-level pathway structures to guide patient-specific multi-view representation learning.
3. **Empirical validation.** We comprehensively evaluate StructAlign across multiple cancer cohorts, demonstrating superior performance and enhanced pathway-level interpretability compared with competitive approaches.

2 Related Work

2.1 Multimodal Cancer Survival Prediction

Multimodal cancer survival prediction aims to quantify prognostic risk by jointly leveraging histopathological slide images and gene expression data (Győrffy, 2021; Ding et al., 2023). Recent approaches organize genes into biologically meaningful groups, such as pathways, where each pathway can be regarded as an individual biological view (Jaume et al., 2024; Yan et al., 2025). This formulation naturally leads to a multi-view learning setting, where each pathway is jointly modeled with the shared histopathological modality to characterize distinct functional aspects of patient risk (Xiong et al., 2024; Qu et al., 2025). However, existing methods mainly focus on learning pathway-specific representations, while overlooking intrinsic relationships among pathways.

In particular, different biological pathways are not independent but exhibit complex inter-pathway relationships, including functional dependencies and heterogeneous prognostic characteristics. Ignoring such inter-pathway structures may limit the learning of biologically meaningful multimodal representations. Different from existing studies, we explore the structural organization among pathways and leverage these structures to guide representation learning.

2.2 Multi-view Learning

Multi-view learning aims to exploit complementary information from multiple view spaces, where each view characterizes different aspects of the same objects (Xu et al., 2024; Yu et al., 2025). In multimodal cancer survival prediction, grouping genes into functional pathways naturally organizes

genomic information into multiple biological views, where each view provides distinct biological guidance and enables the discovery of pathway-specific representations for prediction. Existing pathway-based approaches typically learn pathway-specific representations independently and integrate them through simple fusion strategies like self-attention or weighted aggregation (Zhou & Chen, 2023; Jaume et al., 2024; Zhou et al., 2025).

However, this independent learning-and-fusion paradigm often overlooks the intrinsic relationships among pathway views. As biological pathways form interconnected functional systems, ignoring such structural organization may hinder the learning of biologically consistent representations. In contrast, StructAlign incorporates population-level pathway relationships into patient-specific multimodal learning through structural alignment among pathway views.

3 Exploration of Global Pathway Structure

Pathway-based genomic representation forms a multi-view learning perspective, where each pathway captures a distinct biological process. To characterize global pathway structures, in this section, we perform population-level analyses from two perspectives: **Cross-pathway Correlation** (Sec. 3.1) and **Pathway-survival Association** (Sec. 3.2). We analyze 50 representative human biological pathways from the MSigDB Hallmark gene sets (Liberzon et al., 2015) and present BLCA and UCEC as representative examples. More detailed analyses across all cancer cohorts and pathway information are provided in Appendix A.

3.1 Cross-Pathway Correlation

To characterize relationships among pathway views, we analyze cross-pathway correlations from gene expression profiles. The Gene Set Variation Analysis (GSVA) (Hänzelmann et al., 2013) is used to derive pathway activity scores, forming a pathway activity matrix $\mathbf{S} \in \mathbb{R}^{M \times N}$ for M pathways across N patients. We then quantify the pathway relationships using the Pearson correlation:

$$C_{ij} = \text{Pearson}(\mathbf{S}_i, \mathbf{S}_j), \tag{1}$$

where $C_{ij} \in [-1, 1]$ represents the population-level correlation between pathways i and j. Higher correlations indicate stronger co-activation patterns between these pathways.

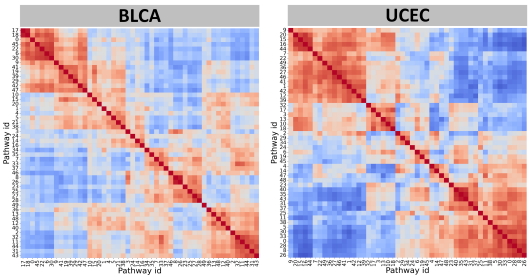

Figure 1: Cross-pathway correlations in BLCA and UCEC cohorts. Rows and columns denote pathway IDs, with red and blue indicating positive and negative correlations, respectively.

The hierarchical clustered correlations are visualized in Fig. 1. Pathway views exhibit structured correlation patterns rather than independent signals, with clustered regions indicating locally organized groups of pathways sharing similar activity patterns across patients. Moreover, the correlation patterns differ substantially across cancer types, reflecting cancer-specific pathway organizations.

3.2 Pathway-Survival Association

Beyond the correlations among pathway views, we further investigate whether different pathways exhibit heterogeneous associations with survival outcomes. Based on these pathway activity scores, we perform univariate Cox regression (Cox, 1972; Therneau, 1997) of each pathway:

$$h(t) = h_0(t) \exp(\beta_m \cdot x_m), \tag{2}$$

where x_m denotes the activity score of pathway m, and β_m is the estimated prognostic coefficient. $\beta_m > 0$ indicates that higher activity is associated with increased risk, whereas $\beta_m < 0$ suggests a protective association with survival.

As shown in Fig. 2, different cancers exhibit distinct pathway-survival associations. In BLCA, Epithelial-Mesenchymal Transition (EMT), Hedgehog signaling, Apical Junction, Myogenesis, and Angiogenesis show positive Cox coefficients, indicating risk-promoting associations.

These pathways are closely associated with tumor progression: EMT promotes tumor cells to acquire migratory and invasive capabilities (Kalluri et al., 2009); angiogenesis facilitates tumor growth by promoting neovascularization and oxygen supply (Folkman, 1995); and Hedgehog signaling contributes to the maintenance of tumor stem cell properties and malignant progression (Taipale & Beachy, 2001). Similarly, UCEC exhibits distinct risk-associated pathways, including IL2/STAT5 signaling, P53, MYC Targets V1, Pancreas Beta Cells, and Peroxisome.

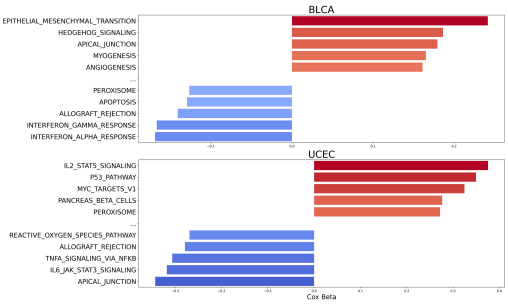

Figure 2: Top-5 risk-promoting and protective pathways ranked by β_m in BLCA and UCEC cohorts. Red and blue indicate positive and negative Cox coefficients, respectively.

Notably, some pathways exhibit cancer-specific prognostic associations; for example, Peroxisome activity is protective in BLCA but risk-promoting in UCEC, highlighting context-dependent pathway functions. These results reveal heterogeneous pathway-survival associations and cancer-specific pathway organizations, motivating adaptive modeling of pathway views for patient-specific survival prediction.

4 The StructAlign Framework

Motivated by these cancer-specific pathway organizations, we propose **StructAlign**, a structural alignment framework integrating pathway-level guidance into patient-specific multi-view learning. All structural guidance is derived exclusively from the training split to avoid information leakage.

As shown in Fig. 3, StructAlign consists of 3 components: **(1) Multi-view Representation Learning** learns pathway-specific representations from multimodal features (Sec.4.2); **(2) Cross-pathway Alignment** preserves global inter-pathway structures (Sec.4.3); and **(3) Pathway-survival Alignment** incorporates survival-associated patterns into patient-specific pathway weighting (Sec.4.4).

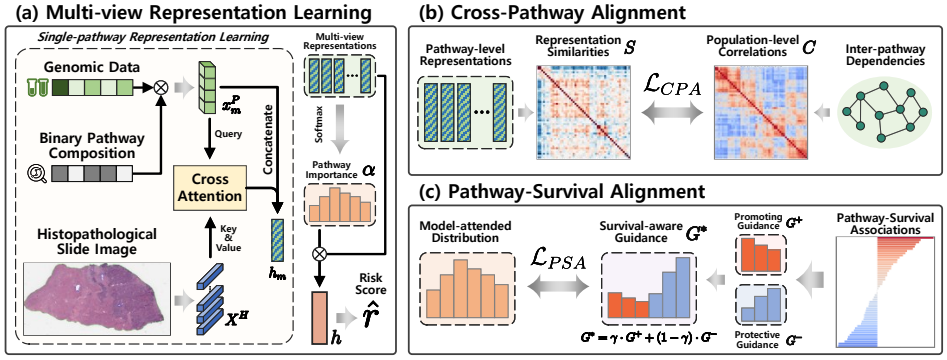

Figure 3: Architectural overview of the proposed StructAlign framework.

4.1 Problem Formulation

For a patient, we denote the histopathological features as $X^H = \{x_l^H\}_{l=1}^L$, where L represents the number of patches extracted from the WSIs. The gene expression profile is organized into M pathway-level views and represented as $X^P = \{x_m^P\}_{m=1}^M$, where x_m^P represents a distinct biological view corresponding to the m-th pathway. All features are preprocessed before model training.

From a multi-view learning perspective, each pathway is treated as an individual biological view with distinct functional characteristics. Each view is jointly modeled with shared histopathological features to learn a pathway-specific multimodal representation h_m via a function $f_m(\cdot)$:

$$\begin{aligned} h_m &= f_m(X^H, x_m^P), \quad m = 1, \ldots, M, \\ \hat{r} &= g(h_1, \ldots, h_M). \end{aligned} \tag{3}$$

The pathway representations are then aggregated by $g(\cdot)$ to estimate the patient-level risk score $\hat{r}$.

4.2 Multi-view Representation Learning

Based on the multi-view formulation, we learn pathway-specific multimodal representations by modeling the interaction between each pathway and pathological features. For m-th pathway, the cross-attention is applied between the pathway feature x_m^P and histopathological features X^H:

$$\tilde{h}_m = \mathrm{CrossAttn}(x_m^P, X^H), \quad h_m = [\tilde{h}_m; x_m^P], \tag{4}$$

where h_m represents the fused representation of the m-th pathway view obtained by concatenating the cross-attended feature $\tilde{h}_m$ with x_m^P. The representations from all pathway views are then aggregated via a softmax weighting operation:

$$\alpha_m = \frac{\exp(\mathrm{FFN}(h_m))}{\sum_{j=1}^{M} \exp(\mathrm{FFN}(h_j))}, \quad h = \sum_{m=1}^{M} \alpha_m h_m, \tag{5}$$

where α_m represents the learned importance of m-th pathway view. The aggregated representation h is then used to estimate the survival risk:

$$\hat{r} = \sigma(\mathrm{FFN}(h)), \quad \mathcal{L}_{NLL} = NLL(\hat{r}), \tag{6}$$

where $\hat{r}$ denotes the predicted risk score, $\mathrm{FFN}(\cdot)$ maps the representation to the risk prediction space, $\sigma(\cdot)$ denotes the sigmoid function, and $\mathcal{L}_{NLL}$ represents the negative log-likelihood loss used for optimization. Detailed computation for $\mathcal{L}_{NLL}$ is provided in Appendix B.

4.3 Cross-Pathway Alignment

Although pathway-specific representations are learned for each patient, different pathways exhibit structured correlation patterns at the population level. To incorporate such global cross-pathway structures into patient-specific representation learning, we align these pathway-level representations with the population-level pathway correlations.

For each patient, we first compute the pairwise cosine similarity between pathway representations:

$$S_{i,j} = \mathrm{CosSim}(h_i, h_j), \tag{7}$$

where $S_{i,j}$ measures the similarity between the representations of the i-th and j-th pathway views. The resulting similarity matrix is denoted as $S \in [-1, 1]^{M \times M}$.

To incorporate the global structural pattern into patient-specific pathway representations, we optimize the consistency between the patient-specific similarity matrix S and the population-level correlation matrix C derived from Eq. 1:

$$\mathcal{L}_{CPA} = \frac{\sum_{i \neq j} (S_{i,j} - C_{i,j})^2}{M(M-1)}. \tag{8}$$

This objective establishes a global-to-instance structural alignment, encouraging patient-level pathway representations to reflect the cross-pathway correlation patterns observed at the population level.

4.4 Pathway-Survival Alignment

Beyond the structured cross-pathway correlations, different pathways exhibit heterogeneous associations with survival outcomes. To incorporate such global pathway-survival structures into patient-specific learning, we first categorize these pathways according to the estimated Cox coefficients β_m. Specifically, we define two pathway sets as:

$$\mathcal{P}^+ = \{m \mid \beta_m > 0\}, \qquad \mathcal{P}^- = \{m \mid \beta_m < 0\}, \tag{9}$$

where $\mathcal{P}^+$ and $\mathcal{P}^-$ denote the sets of **promoting** and **protective** pathways, respectively.

Based on the magnitude of the coefficient β_m, we construct two normalized distributions for the two pathway groups:

$$G_m^+ = \frac{|\beta_m|}{\sum_{j \in \mathcal{P}^+} |\beta_j|}, \quad G_m^- = \frac{|\beta_m|}{\sum_{j \in \mathcal{P}^-} |\beta_j|}, \tag{10}$$

where G_m^+ and G_m^- denote the importance weight of the m-th pathway within the promoting and protective groups. Collectively, these weights form the M-dimensional global pathway importance distributions G^+ and G^-, where pathways outside the corresponding group are assigned zero values.

To provide patient-specific guidance for pathway aggregation, we quantify each patient's relative risk state based on their observed survival time and then define the promoting-pathway weight as:

$$\gamma = \frac{t_{\max} - t}{t_{\max} - t_{\min}}, \tag{11}$$

where t denotes the observed survival time, bounded by $t_{\min}$ and $t_{\max}$ across the cohort. A larger γ corresponds to shorter survival and thus indicates a higher-risk state, whereas a smaller γ reflects longer survival and a lower-risk state. We then combine G^+ and G^- according to this risk-state weight to obtain a patient-specific guidance distribution:

$$G^* = \gamma \cdot G^+ + (1 - \gamma) \cdot G^-. \tag{12}$$

Finally, we align the patient-specific pathway importance distribution α (from the softmax aggregation in Eq. 5) with the guidance distribution G^* using the Kullback–Leibler divergence:

$$\mathcal{L}_{PSA} = \sum\nolimits_{m=1}^{M} G_m^* \cdot \log \frac{G_m^*}{\alpha_m}. \tag{13}$$

By minimizing $\mathcal{L}_{PSA}$, the learned pathway importance distribution is guided toward the survival-associated pseudo-label distribution G^*, encouraging the model to assign greater attention to promoting pathways for shorter survival and to protective pathways for longer survival.

4.5 Overall Optimization

The StructAlign framework jointly optimizes the survival prediction objective and the two complementary structural alignment objectives. The overall training objective is formulated as:

$$\mathcal{L} = \mathcal{L}_{NLL} + \lambda_1 \cdot \mathcal{L}_{CPA} + \lambda_2 \cdot \mathcal{L}_{PSA}, \tag{14}$$

where λ_1 and λ_2 are hyper-parameters controlling the contributions of the two structural alignment objectives. In contrast to the CPA objective, which is applied to all patients, the PSA objective is applied only to patients with events, as censored patients lack explicit event times.

By jointly optimizing these objectives, StructAlign learns patient-specific pathway representations that are not only predictive of survival outcomes but also structurally consistent with global cross-pathway correlations and pathway-outcome associations.

5 Experiments

5.1 Benchmark Settings

To comprehensively evaluate StructAlign, we conduct experiments on six cancer cohorts from TCGA, including BLCA, BRCA, CESC, GBMLGG, LUAD, and UCEC. We compare StructAlign with representative baselines, including MCAT (Chen et al., 2021), MOTCat (Xu & Chen, 2023), CMTA (Zhou & Chen, 2023), SurvPath (Jaume et al., 2024), PIBD (Zhang et al., 2024), and LD-CVAE (Zhou et al., 2025).

We use the concordance index (**C-index**) with 5-fold cross-validation for evaluation. The loss hyper-parameters in Eq. (14) are empirically set to $\lambda_1 = 0.5$ and $\lambda_2 = 0.5$. Detailed C-index computation, implementation details, and hyperparameter analysis are provided in Appendix B.

5.2 Main Results

We evaluate StructAlign against a diverse set of baseline models on multiple cancer cohorts, with the corresponding C-index results summarized in Table 1. Across the evaluated datasets, StructAlign achieves the best overall performance, consistently surpassing existing competing methods. Notably, it improves the C-index by an average of 3.9% over the strongest baseline, LD-CVAE.

Table 1: Overall C-index comparison across six cancer cohorts, with N denoting the number of patients. **Bold** values indicate the best and underline values indicate the runner-up. Results are based on 5-fold cross-validation and reported as mean ± standard deviation (std).

Method	BLCA ($N = 373$)	BRCA ($N = 957$)	CESC ($N = 241$)	GBMLGG ($N = 571$)	LUAD ($N = 452$)	UCEC ($N = 480$)
MCAT	0.630±0.034	0.625±0.039	0.705±0.055	0.812±0.036	0.629±0.059	0.694±0.023
MOTCat	0.633±0.018	0.650±0.007	0.674±0.062	0.821±0.038	0.621±0.061	0.680±0.026
CMTA	0.632±0.032	0.612±0.027	0.721±0.038	0.816±0.035	0.610±0.046	0.656±0.049
SurvPath	0.615±0.031	0.638±0.030	0.693±0.061	0.821±0.037	0.620±0.056	0.704±0.032
PIBD	0.618±0.020	0.627±0.028	0.719±0.058	0.822±0.021	0.637±0.031	0.703±0.027
LD-CVAE	0.641±0.061	0.637±0.022	0.715±0.021	0.812±0.034	0.624±0.029	0.707±0.028
StructAlign	**0.683±0.037**	**0.674±0.081**	**0.741±0.064**	**0.859±0.027**	**0.672±0.064**	**0.742±0.034**

The consistent improvements across cancer types can be attributed to StructAlign's global-to-instance structural alignment strategy, which incorporates population-level pathway structures into patient-specific representation learning. The cross-pathway alignment preserves inter-pathway correlations, while the pathway-survival alignment captures heterogeneous prognostic roles by guiding pathway importance toward survival-associated patterns.

By jointly modeling these complementary structures, StructAlign enables effective multi-view learning beyond simple pathway fusion, capturing both pathway-specific information and intrinsic correlations among pathway views to learn predictive and structurally consistent representations.

Table 2: Ablation evaluation of different alignment configurations in StructAlign. CPA and PSA denote the Cross-Pathway and Pathway-Survival Alignment objectives, respectively.

Alignment		Cancer Types					
CPA	PSA	BLCA	BRCA	CESC	GBMLGG	LUAD	UCEC
✓	✓	**0.683±0.037**	**0.674±0.081**	**0.741±0.064**	**0.859±0.027**	**0.672±0.064**	**0.742±0.034**
✓	✗	0.646±0.030	0.668±0.084	0.737±0.073	0.851±0.029	0.645±0.066	0.710±0.051
✗	✓	0.649±0.039	0.659±0.054	0.715±0.050	0.855±0.027	0.649±0.055	0.707±0.034
✗	✗	0.619±0.025	0.630±0.047	0.690±0.062	0.821±0.036	0.629±0.045	0.688±0.038

5.3 Ablation Study

To investigate the contribution of each alignment component in StructAlign, we conduct ablation studies by removing the **CPA** and **PSA** objectives. The results are summarized in Table 2, StructAlign achieves the best performance across all cancer cohorts when both alignment objectives are incorporated, demonstrating the effectiveness of the proposed structural constraints. Removing CPA and PSA results in average C-index decreases of 2.3% and 1.9%, respectively.

The degradation after removing CPA indicates that modeling inter-pathway correlations helps preserve the intrinsic structural organization among pathway views, while the performance drop caused by removing PSA highlights the importance of capturing heterogeneous pathway-outcome associations for patient-specific risk prediction. When both alignment objectives are removed, the average C-index further decreases, confirming that CPA and PSA provide complementary structural guidance and that incorporating population-level pathway structures is essential for learning predictive and biologically consistent multimodal representations.

5.4 Structural Consistency of Representations

To assess whether StructAlign preserves intrinsic relationships among pathway views, we compute the pairwise similarity matrix of the learned pathway representations h_m for each patient and average these matrices across patients. These similarity matrices are compared with the population-level pathway correlation matrices from Sec. 3.1 to evaluate whether the learned views retain the underlying biological organizations among pathway views. Additional results on other cancer types are provided in Appendix A.

As shown in Fig. 4, the similarity patterns among learned pathway views closely align with the population-level pathway correlations, indicating that StructAlign effectively preserves inter-view correlations during patient-specific multi-view representation learning. These results demonstrate that the proposed cross-pathway alignment successfully transfers cohort-level biological structures into learned pathway representations, enabling structure-aware multi-view learning.

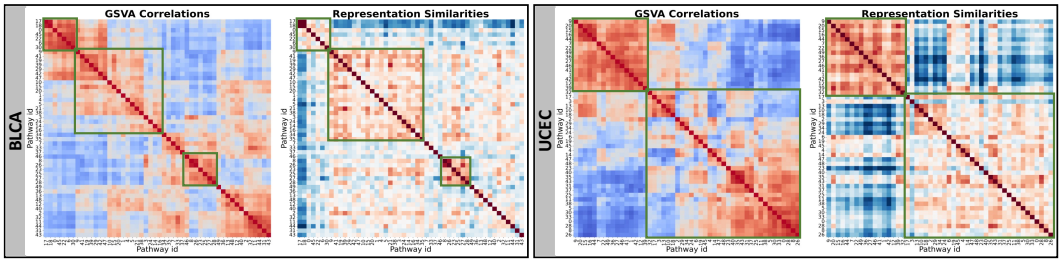

Figure 4: Pathway representation similarities versus pathway correlation structures. For a fair comparison, the hierarchical clustering order derived from the original pathway correlation matrices is retained for both matrices. Green boxes highlight regions with consistent structural patterns.

5.5 Survival-associated Pathway View Preferences

To investigate patient-specific pathway preferences learned by StructAlign, we analyze the importance α in Eq. (5) across samples. Specifically, we examine the aggregated importance assigned to 'promoting' and 'protective' pathway groups. Samples are ordered according to their observed event times, and only uncensored patients are included.

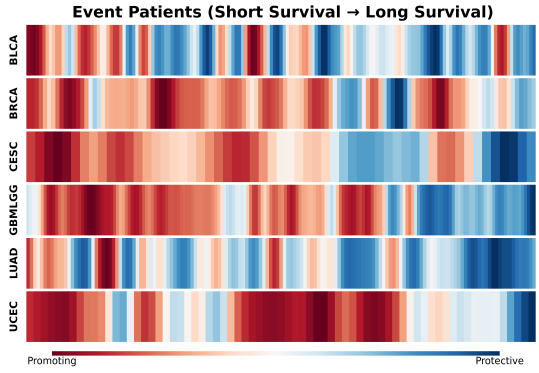

Figure 5: Patient-specific view preferences between promoting and protective pathways.

As shown in Fig. 5, patients with shorter survival times tend to assign higher importance to promoting pathways, whereas patients with longer survival times exhibit greater attention toward protective pathways. This trend indicates that StructAlign learns clinically meaningful pathway preferences and adapts pathway importance according to patient-specific survival characteristics and demonstrates the effectiveness of the pathway-survival alignment in capturing heterogeneous prognostic patterns among pathway views.

5.6 Kaplan-Meier Overall Survival Analysis

Beyond the C-index evaluation, we further assess the clinical relevance of StructAlign by stratifying patients into two equal-sized groups, denoted as 'high-risk' and 'low-risk', according to their predicted risk scores across different cancer cohorts. The overall survival differences between the two groups are then evaluated using the Kaplan-Meier (KM) analysis (Kaplan & Meier, 1958), with the results compared against those of the strongest baseline, LD-CVAE.

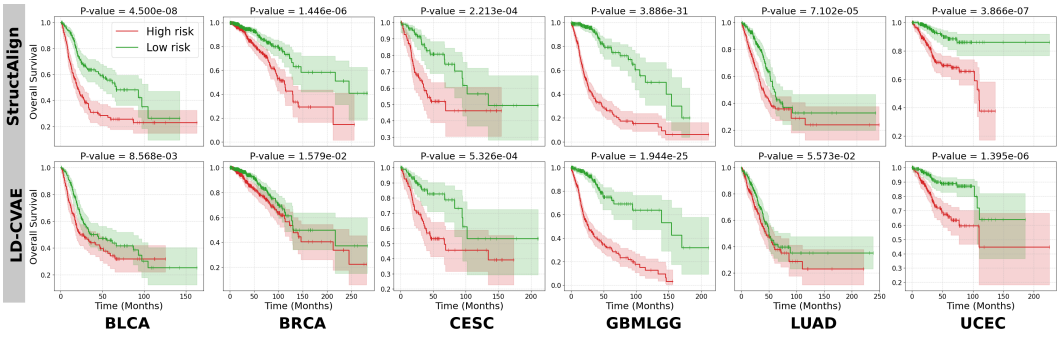

Figure 6: Kaplan-Meier analysis of high-risk and low-risk groups on StructAlign and LD-CVAE.

The results and corresponding p-values from the Log-rank test (Mantel et al., 1966) are presented in Fig. 6. StructAlign consistently achieves clearer separation between high-risk and low-risk groups than LD-CVAE across different cancer cohorts, with more significant survival differences between the two groups. These results indicate that StructAlign not only improves risk prediction accuracy but also captures clinically meaningful prognostic patterns, enabling more reliable patient stratification based on survival outcomes.

5.7 Attribution Visualization

To interpret the pathway-specific prediction of StructAlign, we visualize the cross-attention weights from pathway representations onto histopathological regions. The pathological interpretations were reviewed by experienced pathologists. As shown in Fig. 7, we present two representative BLCA cases from the high-risk and low-risk groups defined in Sec. 5.6. Additional attribution visualizations for other cancer cohorts are provided in Appendix C.

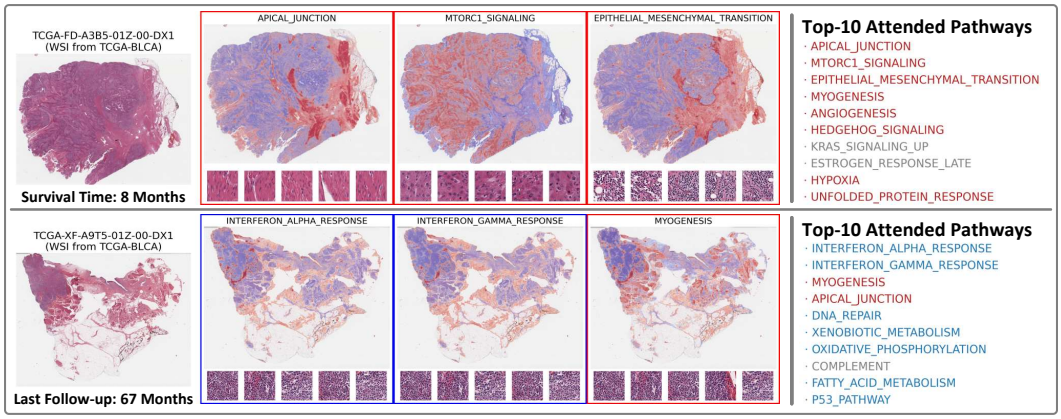

Figure 7: Visualization of cross-attention weights in StructAlign. We visualize the three pathways with the highest learned importance, where each attention map highlights the top-5 patches with the highest attention scores. Red, blue, and gray indicate the top-15 promoting pathways, top-15 protective pathways, and remaining pathways, respectively.

For the high-risk case, the EMT pathway highlights invasive tumor regions with stromal infiltration, while the Apical Junction and mTORC1 signaling pathways focus on glandular structures and highly proliferative regions with increased nuclear density, respectively. In contrast, the low-risk case shows strong attribution for the INTERFERON ALPHA RESPONSE and INTERFERON GAMMA RESPONSE pathways in regions with preserved tissue architecture. These observations suggest that StructAlign effectively captures pathway-specific spatial patterns and provides interpretable associations between biological pathway activities and histopathological morphology.

6 Conclusion

In this work, we investigate previously overlooked global structures among biological pathways from a population-level perspective, revealing intrinsic cross-pathway correlations and heterogeneous pathway-survival associations across cancers. Motivated by these observations, we propose StructAlign, a global-to-instance structural alignment framework for multimodal cancer survival prediction. By incorporating these pathway-level structures, StructAlign introduces complementary alignment objectives to guide structurally consistent pathway-based multi-view learning.

Extensive experiments across six TCGA cancer cohorts demonstrate that StructAlign consistently outperforms state-of-the-art methods. Ablation and visualization analyses further validate the complementary effects of the proposed alignment objectives and their ability to preserve meaningful pathway-level structures. Overall, these results highlight the importance of integrating global biological pathway structures into patient-specific representation learning and demonstrate the potential of structural alignment for biologically informed multi-view learning.

AI USE STATEMENT

In this work, we used generative AI tools for language polishing and improving the clarity and readability of the manuscript. We did not use generative AI tools for information retrieval or discovery, research ideation, research execution, drafting substantive sections of the paper, generating synthetic datasets, or proving mathematical claims. We have reviewed all AI-assisted text and verified its accuracy and consistency with the research conducted by the authors. The authors take full responsibility for the final content of this work, including all text, claims, results, and other artifacts produced with the assistance of generative AI tools.

REFERENCES

Martim Afonso, Praphulla MS Bhawsar, Monjoy Saha, Jonas S Almeida, and Arlindo L Oliveira. Multiple instance learning for wsi: A comparative analysis of attention-based approaches. *Journal of Pathology Informatics*, 15:100403, 2024.

Ethan Cerami, Jianjiong Gao, Ugur Dogrusoz, Benjamin E Gross, Selcuk Onur Sumer, Bülent Arman Aksoy, Anders Jacobsen, Caitlin J Byrne, Michael L Heuer, Erik Larsson, et al. The cbio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer discovery*, 2(5):401–404, 2012.

Richard J Chen, Ming Y Lu, Wei-Hung Weng, Tiffany Y Chen, Drew FK Williamson, Trevor Manz, Maha Shady, and Faisal Mahmood. Multimodal co-attention transformer for survival prediction in gigapixel whole slide images. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 4015–4025, 2021.

David R Cox. Regression models and life-tables. *Journal of the royal statistical society: Series B (methodological)*, 34(2):187–202, 1972.

Kexin Ding, Mu Zhou, Dimitris N Metaxas, and Shaoting Zhang. Pathology-and-genomics multimodal transformer for survival outcome prediction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 622–631. Springer, 2023.

Omar SM El Nahhas, Chiara ML Loeffler, Zunamys I Carrero, Marko van Treeck, Fiona R Kolbinger, Katherine J Hewitt, Hannah S Muti, Mara Graziani, Qinghe Zeng, Julien Calderaro, et al. Regression-based deep-learning predicts molecular biomarkers from pathology slides. *nature communications*, 15(1):1253, 2024.

Judah Folkman. Angiogenesis in cancer, vascular, rheumatoid and other disease. *Nature medicine*, 1(1):27–30, 1995.

Balázs Győrffy. Survival analysis across the entire transcriptome identifies biomarkers with the highest prognostic power in breast cancer. *Computational and structural biotechnology journal*, 19:4101–4109, 2021.

Douglas Hanahan and Robert A Weinberg. Hallmarks of cancer: the next generation. *cell*, 144(5): 646–674, 2011.

Sonja Hänzelmann, Robert Castelo, and Justin Guinney. Gsva: gene set variation analysis for microarray and rna-seq data. *BMC bioinformatics*, 14(1):7, 2013.

Frank E Harrell Jr, Kerry L Lee, and Daniel B Mark. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in medicine*, 15(4):361–387, 1996.

Guillaume Jaume, Anurag Vaidya, Richard J Chen, Drew FK Williamson, Paul Pu Liang, and Faisal Mahmood. Modeling dense multimodal interactions between biological pathways and histology for survival prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11579–11590, 2024.

Raghu Kalluri, Robert A Weinberg, et al. The basics of epithelial-mesenchymal transition. *The Journal of clinical investigation*, 119(6):1420–1428, 2009.

Edward L Kaplan and Paul Meier. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481, 1958.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Arthur Liberzon, Chet Birger, Helga Thorvaldsdóttir, Mahmoud Ghandi, Jill P Mesirov, and Pablo Tamayo. The molecular signatures database hallmark gene set collection. *Cell systems*, 1(6): 417–425, 2015.

Ming Y Lu, Drew FK Williamson, Tiffany Y Chen, Richard J Chen, Matteo Barbieri, and Faisal Mahmood. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nature biomedical engineering*, 5(6):555–570, 2021.

Nathan Mantel et al. Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother Rep*, 50(3):163–170, 1966.

Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. In *NIPS 2017 Workshop on Autodiff*, 2017.

Mingcheng Qu, Guang Yang, Donglin Di, Yue Gao, Tonghua Su, Yang Song, and Lei Fan. Memory-augmented incomplete multimodal survival prediction via cross-slide and gene-attentive hypergraph learning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 318–327. Springer, 2025.

Francisco Sanchez-Vega, Marco Mina, Joshua Armenia, Walid K Chatila, Augustin Luna, Konnor C La, Sofia Dimitriadoy, David L Liu, Havish S Kantheti, Sadegh Saghafinia, et al. Oncogenic signaling pathways in the cancer genome atlas. *Cell*, 173(2):321–337, 2018.

Andrew H Song, Guillaume Jaume, Drew FK Williamson, Ming Y Lu, Anurag Vaidya, Tiffany R Miller, and Faisal Mahmood. Artificial intelligence for digital and computational pathology. *Nature Reviews Bioengineering*, 1(12):930–949, 2023.

Jussi Taipale and Philip A Beachy. The hedgehog and wnt signalling pathways in cancer. *nature*, 411(6835):349–354, 2001.

Terry M Therneau. Extending the cox model. In *Proceedings of the first Seattle symposium in biostatistics: survival analysis*, pp. 51–84. Springer, 1997.

Xiyue Wang, Sen Yang, Jun Zhang, Minghui Wang, Jing Zhang, Junzhou Huang, Wei Yang, and Xiao Han. Transpath: Transformer-based self-supervised learning for histopathological image classification. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 186–195. Springer, 2021.

Xiyue Wang, Sen Yang, Jun Zhang, Minghui Wang, Jing Zhang, Wei Yang, Junzhou Huang, and Xiao Han. Transformer-based unsupervised contrastive learning for histopathological image classification. *Medical image analysis*, 81:102559, 2022.

John N Weinstein, Eric A Collisson, Gordon B Mills, Kenna R Shaw, Brad A Ozenberger, Kyle Ellrott, Ilya Shmulevich, Chris Sander, and Joshua M Stuart. The cancer genome atlas pan-cancer analysis project. *Nature genetics*, 45(10):1113–1120, 2013.

Conghao Xiong, Hao Chen, Hao Zheng, Dong Wei, Yefeng Zheng, Joseph JY Sung, and Irwin King. Mome: Mixture of multimodal experts for cancer survival prediction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 318–328. Springer, 2024.

Cai Xu, Jiajun Si, Ziyu Guan, Wei Zhao, Yue Wu, and Xiyue Gao. Reliable conflictive multi-view learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pp. 16129–16137, 2024.

Yingxue Xu and Hao Chen. Multimodal optimal transport-based co-attention transformer with global structure consistency for survival prediction. In *2023 IEEE/CVF international conference on computer vision (ICCV)*, pp. 21184–21194. IEEE, 2023.

Rui Yan, Xueyuan Zhang, Zihang Jiang, Baizhi Wang, Xiuwu Bian, Fei Ren, and S Kevin Zhou. Pathway-aware multimodal transformer (pamt): Integrating pathological image and gene expression for interpretable cancer survival analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025.

Zhiwen Yu, Ziyang Dong, Chenchen Yu, Kaixiang Yang, Ziwei Fan, and CL Philip Chen. A review on multi-view learning. *Frontiers of Computer Science*, 19(7):197334, 2025.

Shekoufeh Gorgi Zadeh and Matthias Schmid. Bias in cross-entropy-based training of deep survival networks. *IEEE transactions on pattern analysis and machine intelligence*, 43(9):3126–3137, 2020.

Tian-mei Zeng, Guang Yang, Cheng Lou, Wei Wei, Chen-jie Tao, Xi-yun Chen, Qin Han, Zhuo Cheng, Pei-pei Shang, Yu-long Dong, et al. Clinical and biomarker analyses of sintilimab plus gemcitabine and cisplatin as first-line treatment for patients with advanced biliary tract cancer. *Nature communications*, 14(1):1340, 2023.

Yilan Zhang, Yingxue Xu, Jianqi Chen, Fengying Xie, and Hao Chen. Prototypical information bottlenecking and disentangling for multimodal cancer survival prediction. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=otHZ8JAIgh.

Fengtao Zhou and Hao Chen. Cross-modal translation and alignment for survival analysis. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 21485–21494, 2023.

Junjie Zhou, Jiao Tang, Yingli Zuo, Peng Wan, Daoqiang Zhang, and Wei Shao. Robust multimodal survival prediction with conditional latent differentiation variational autoencoder. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10384–10393, 2025.

A Global Pathway Structure Analysis

A.1 GSVA Implementation

The GSVA-based analysis is implemented in R (version 4.5.2) using the GSVA package. Given the pathway activity matrix $\mathbf{S} = [s_{ij}] \in \mathbb{R}^{M \times N}$, where s_{ij} denotes the activity score of the j-th pathway for the i-th patient, we further perform Z-score normalization across patients for each pathway:

$$\hat{s}_{ij} = \frac{s_{ij} - \mu_j}{\sigma_j},$$

where μ_j and σ_j represent the mean and standard deviation of the j-th pathway activity across the cohort, respectively. This normalization ensures comparable scales among pathway views and enables fair estimation of pathway-level associations with survival outcomes.

In the analysis stage, the pathway activity matrix is constructed using the entire cohort. During training, the structural guidance is generated only from the training set to avoid information leakage.

A.2 Analysis on Cross-Pathway Structures

To characterize the structural organization among pathway views, we analyze the cross-pathway correlations based on the GSVA-derived activity matrix. Specifically, we calculate the Pearson correlation between each pair of pathways to construct a cohort-level pathway correlation matrix. The hierarchical clustering is applied to reveal structural patterns.

As shown in Fig. 8, different cohorts exhibit heterogeneous correlation patterns, suggesting that pathway relationships vary across cancer types. Despite such heterogeneity, the correlation distributions across cohorts consistently span a broad range (about -0.75 to 0.75), with median values close to zero. This indicates that pathway views are neither fully independent nor uniformly correlated, but instead exhibit structured dependencies with both positive and negative associations, demonstrating the existence of non-trivial global organizations among pathway views and motivate the incorporation of population-level structural guidance for learning.

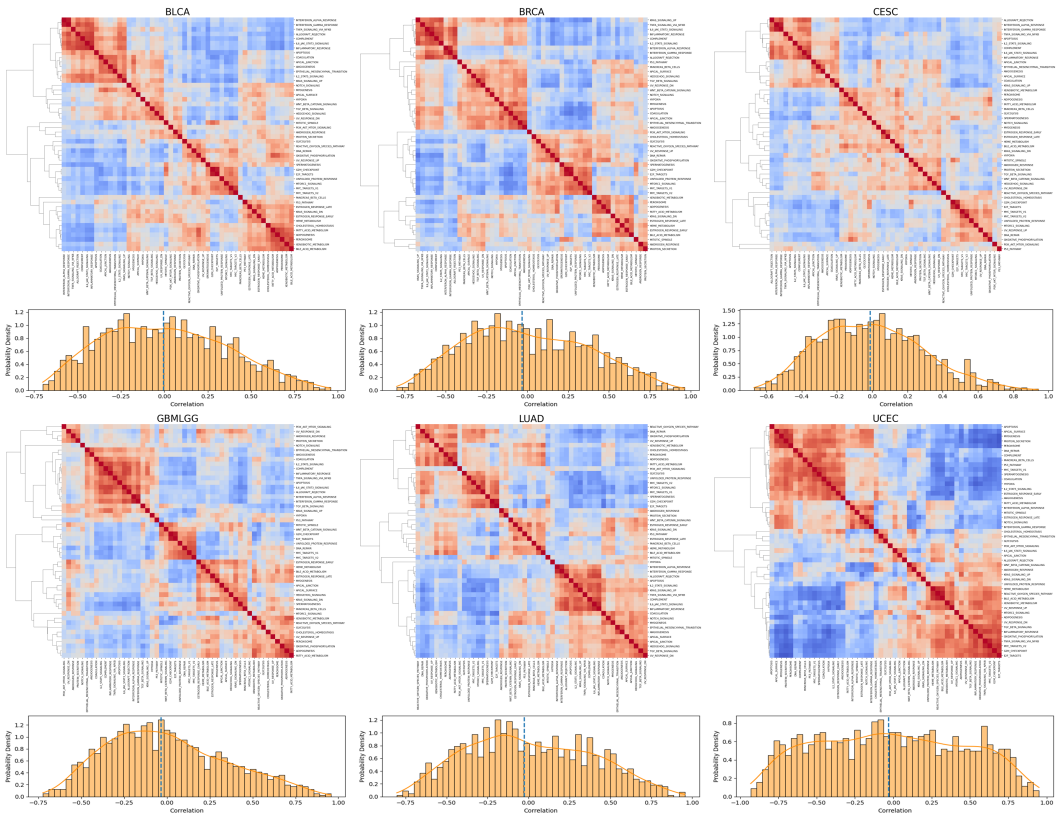

Figure 8: Cross-pathway correlations across cancer cohorts. For each cohort, the upper panel shows the hierarchical clustering heatmap correlation. Rows and columns denote pathway IDs, with red and blue indicating positive and negative correlations, respectively. The lower panel presents the corresponding distribution, with blue dashed lines indicating the median value.

A.3 Analysis on Pathway-Survival Structures

To characterize the associations between pathway activities and patient survival outcomes, we perform univariate Cox regression analysis for each pathway based on the GSVA-derived pathway activity scores. As shown in Fig. 9, different cancer cohorts exhibit heterogeneous pathway-survival associations, the estimated Cox coefficients reveal distinct prognostic patterns among pathways, where some pathways consistently exhibit risk-promoting associations while others are associated with protective associations across cancer types. These observations suggest that pathway views contain distinct prognostic characteristics and should not be considered equally informative for patient-specific survival prediction.

By analyzing pathways with the top-5 highest and lowest β values, we further identify both shared and cancer-specific prognostic patterns across cohorts. Several pathways exhibit consistent tendencies across multiple cancer types. For example, *EPITHELIAL_MESENCHYMAL_TRANSITION* is identified as a promoting pathway in BLCA, CESC, and GBMLGG, while *ALLOGRAFT_REJECTION* shows protective associations in BLCA, BRCA, CESC, and UCEC. Moreover, *GLYCOLYSIS* consistently appears as a promoting pathway in BRCA, GBMLGG, and LUAD.

Meanwhile, some demonstrate cancer-dependent prognostic behaviors. *INTERFERON_ALPHA_RESPONSE* and *INTERFERON_GAMMA_RESPONSE* are associated with increased risk in GBMLGG but exhibit protective associations in BLCA. Similarly, *PANCREAS_BETA_CELLS* shows a promoting tendency in UCEC but protective associations in GBMLGG and LUAD. These observations reveal that pathway-survival relationships contain both shared biological patterns and cancer-specific characteristics.

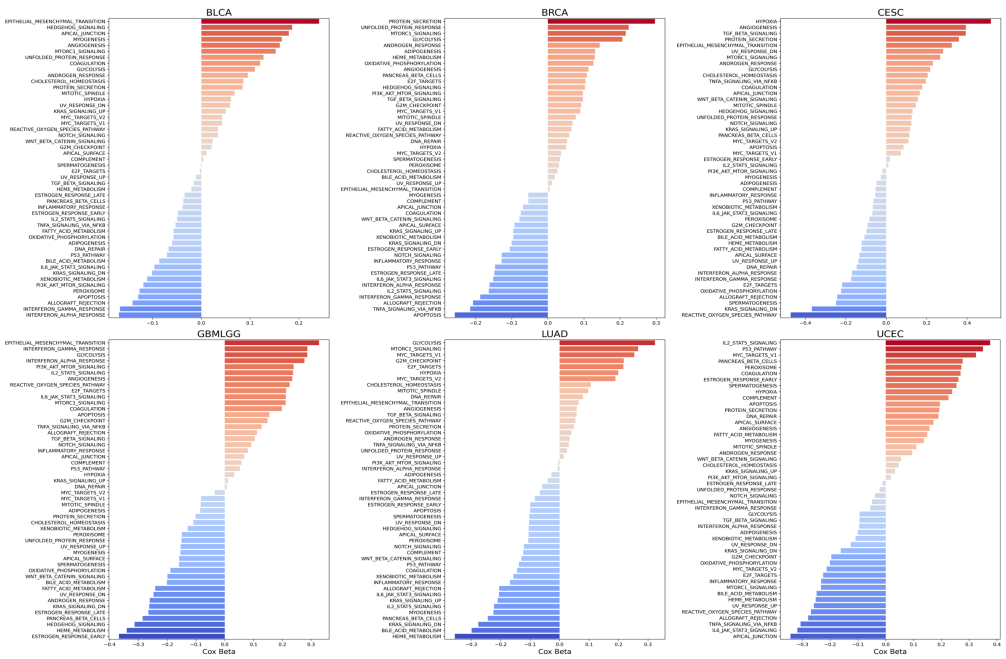

Figure 9: Pathway-survival associations across cancer cohorts. The heatmap shows Cox coefficients β estimated from Z-score normalized pathway activity scores. Positive and negative coefficients indicate promoting and protective pathways, respectively.

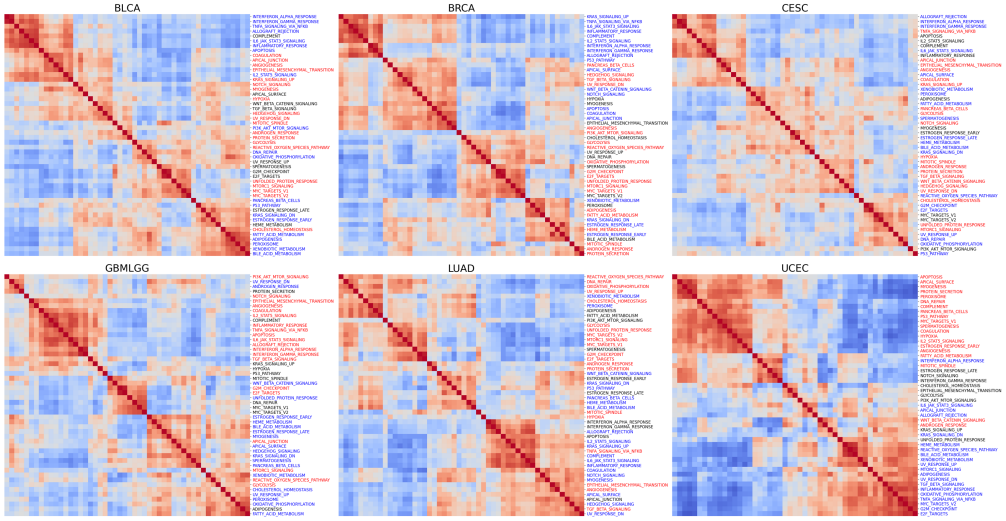

Figure 10: Relationship between pathway-survival associations and clustered cross-pathway structures. Pathways are colored according to their Cox coefficients, where red and blue indicate the top-20 promoting and protective pathways, respectively, and black denotes the remaining pathways.

A.4 Global Structure of Prognostic Pathways

To investigate whether pathway-survival associations are reflected in the global pathway structures, we analyze the relationship between Cox coefficients and the hierarchical clustering results of pathway correlations. As shown in Fig. 10, pathways with similar prognostic tendencies tend to form localized structural clusters across different cancer cohorts, suggesting that pathways with related survival associations often exhibit coordinated biological relationships.

We observe that prognostically consistent pathways are frequently grouped within the same correlation-based clusters, particularly in GBMLGG, LUAD, and UCEC. Notably, in the UCEC cohort, risk-promoting and protective pathways demonstrate a clear separation tendency, forming distinct structural groups with different prognostic characteristics.

These observations indicate that pathway-survival associations are not independent of pathway interactions but are embedded within the global organization of pathway views. Therefore, incorporating such population-level structures provides valuable guidance for multi-view representation learning by aligning learned pathway importance with biologically meaningful organizations.

A.5 Structural Consistency across All Cohorts

To further validate the structural preservation of StructAlign, we extend the analysis to all six cancer cohorts. The learned pathway representations h_m are used to construct patient-level pathway similarity matrices, which are compared with the population-level pathway correlation structures.

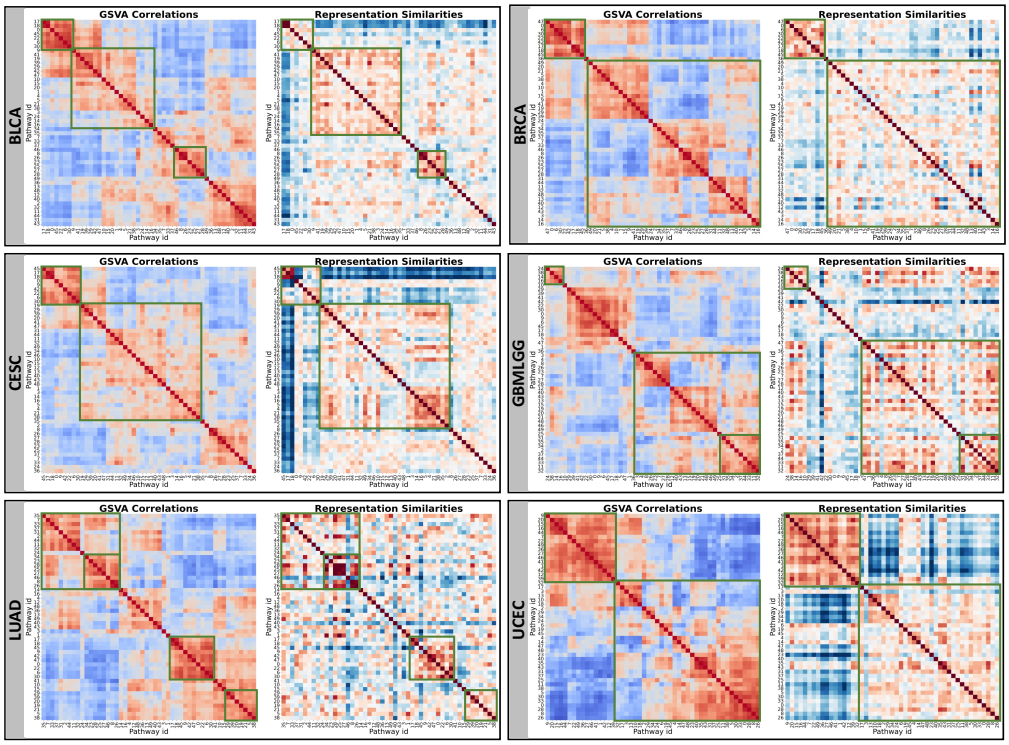

Figure 11: Visualization of pathway representation similarities versus pathway correlation structures across six cancer cohorts. The hierarchical clustering order derived from the pathway correlation matrices is retained for both matrices. Green boxes highlight regions with consistent structural patterns between learned representations and pathway correlations.

As shown in Fig. 11, StructAlign preserves consistent structural patterns across most cancer cohorts. BLCA, CESC, GBMLGG, and UCEC exhibit clear regional correspondence between learned representation similarities and pathway correlations, indicating that the model effectively captures pathway structures. For BRCA, the consistency is mainly observed within localized pathway regions, while the remaining pathways still exhibit relatively weak but discernible structural correspondence across a broader range of pathway views. Although LUAD presents more heterogeneous similarity patterns with weaker overall correspondence, several distinct regions still demonstrate structural consistency.

These results confirm that the proposed cross-pathway alignment can effectively incorporate population-level pathway structures into patient-specific representation learning while adapting to cancer-specific pathway organizations.

B Implementation Details

B.1 Genomic Data Implementation

We select 50 human biological pathways from the MSigDB Hallmark gene sets (Liberzon et al., 2015), detailed list of these pathways is provided in the Supplementary Materials. The genomic information comprises mutation (MUT) profiles, copy number variations (CNV), and RNA sequencing data. These data are publicly accessible via the cBioPortal database (Cerami et al., 2012).

B.2 Data Pre-processing

We perform following preprocessing steps before training.

Histopathology features. To enable efficient computation, each WSI is divided into multiple patches. Each patch is encoded into a feature vector using a fixed pretrained encoder. Specifically, we employ the histopathology foundation model CTransPath (Wang et al., 2021; 2022) by default to extract patch-level representations. This process transforms each WSI into a histopathological feature matrix $X^H \in \mathbb{R}^{L\times d}$, where L denotes the number of patches and d represents the patch feature dimension.

Genomic features. The gene expression profile is grouped into M biologically informed pathways, with the genes belonging to the m-th pathway represented as $\tilde{x}_m^P$. Since different pathways contain varying numbers of genes, we apply a trainable pathway-specific encoder Ω_m (a linear layer by default) to project each pathway feature into a shared embedding space of dimension d. The resulting pathway-level genomic representation is denoted as $X^P \in \mathbb{R}^{M\times d}$.

B.3 C-index Computation

The concordance index (C-index) (Harrell Jr et al., 1996) measures the agreement between predicted risk scores and the observed ordering of patient survival outcomes. Unlike metrics that evaluate absolute prediction errors, the C-index assesses whether the relative risk ranking between patients is consistent with their survival outcomes.

For each patient, T denotes the observed survival time or the last follow-up time, and c denotes the censorship indicator, where $c = 0$ indicates that the event (death) is observed and $c = 1$ indicates a censored observation. For deceased patients, T represents the survival time from the initial recording time to the observed event. For a pair of patients (i, j), the pair is considered **comparable** if the survival ordering can be determined, i.e.,

1. The event is observed for patient i ($c_i = 0$); and
2. Patient i has a shorter survival time than j ($T_i < T_j$).

A comparable pair is considered **concordant** if the patient with higher predicted risk score has shorter survival time:

$$\hat{r}_i > \hat{r}_j \quad \text{if} \quad T_i < T_j.$$

The C-index is calculated as the proportion of concordant pairs among all comparable pairs:

$$\text{C-index} = \frac{\text{concordant}}{\text{comparable}}.$$

B.4 Negative Log-likelihood Loss

For time-aware modeling and stable optimization, we follow prior studies (Jaume et al., 2024; Zhang et al., 2024; Zhou et al., 2025) to divide the overall observation period into four quartile-based time intervals, denoted as $t_1, \ldots, t_4$. The model estimates the event probability within each interval, producing interval-wise risk predictions $\left(\hat{r}^{(t_1)}, \ldots, \hat{r}^{(t_4)}\right)$ for each patient. For inference, these interval-wise predictions are averaged to obtain a unified patient-specific risk score: $\hat{r} = \frac{1}{4}\sum_{j=1}^{4} \hat{r}^{(t_j)}$. We

adopt the NLL loss (Zadeh & Schmid, 2020) for discrete-time survival optimization. For each patient, the survival time is assigned to a corresponding time interval label $\tau \in \{1, 2, 3, 4\}$:

$$\tau = j, \quad \text{if } t_{j-1} < T \leq t_j \, .$$

The cumulative survival probability up to the j-th interval is computed as: $\hat{s}^{(t_j)} = \prod_{k=1}^{j}(1 - \hat{r}^{(t_k)})$. The patient-specific NLL loss is formulated as:

$$\mathcal{L}_{NLL} = -c \cdot \log\left(\hat{s}^{(t_\tau)}\right) - (1-c) \cdot \left(\log\left(\hat{s}^{(t_{\tau-1})}\right) + \log\left(\hat{r}^{(t_\tau)}\right)\right).$$

For patients with events ($c = 0$), the loss models survival before and event occurrence within the interval. For censored patients ($c = 1$), it only models survival up to the last observed interval.

B.5 Experimental Settings

All models are implemented in PyTorch (Paszke et al., 2017) and trained on a single NVIDIA A800 GPU, with the Adam optimizer (Kingma & Ba, 2014) adopted. Due to the large-scale pathology features, we use a batch size of 1 with gradient accumulation over 16 iterations. For pathology pre-processing, WSIs are locally tiled into 512×512 patches following the implementations of CLAM (Lu et al., 2021) and CTransPath (Wang et al., 2022). To ensure absolute fairness, all baselines and StructAlign were rigorously evaluated using the exactly same locally pre-processed features and cross-validation splits.

B.6 Hyperparameter Sensitivity Analysis

We analyze the sensitivity of two key hyperparameters, λ_1 and λ_2, which control the strengths of the cross-pathway alignment (CPA) and pathway-survival alignment (PSA), respectively. Each parameter is varied from 0.1 to 1.0 while fixing the other to 0.5. The results are presented in Fig. 12.

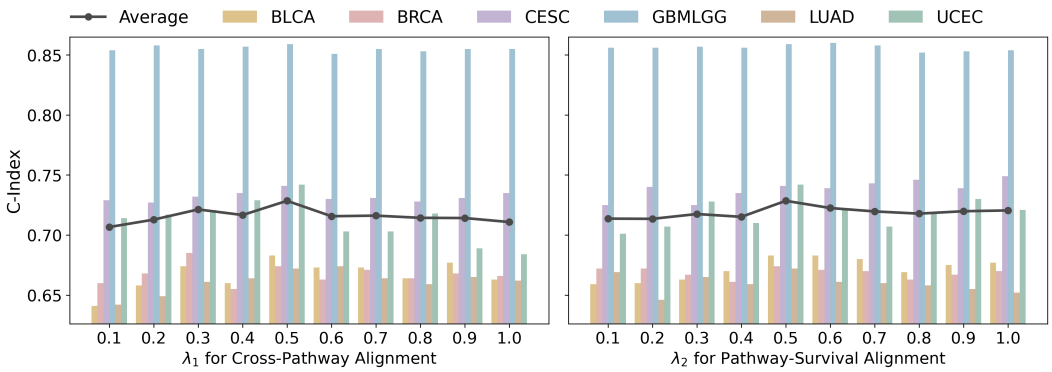

Figure 12: Sensitivity analysis of the alignment weights λ_1 and λ_2, which control the strengths of Cross-Pathway Alignment (CPA) and Pathway-Survival Alignment (PSA). During the analysis of one hyperparameter, the other is fixed to 0.5 by default.

StructAlign maintains relatively stable performance across different configurations, demonstrating the robustness of the proposed alignment objectives. This stability suggests that the model can effectively benefit from both structural constraints without being overly sensitive to their weighting. Although the optimal configuration varies among cancer cohorts due to differences in biological characteristics, the overall average C-index achieves the best performance when both λ_1 and λ_2 are set to 0.5. Therefore, we adopt $\lambda_1 = \lambda_2 = 0.5$ as the default setting in all experiments.

C Attribution Analysis

We further analyze pathway-specific attention distributions across all cancer cohorts. This analysis aims to examine whether the learned pathway representations capture biologically meaningful morphological patterns and whether pathway-level contributions are consistent with the identified survival-associated characteristics.

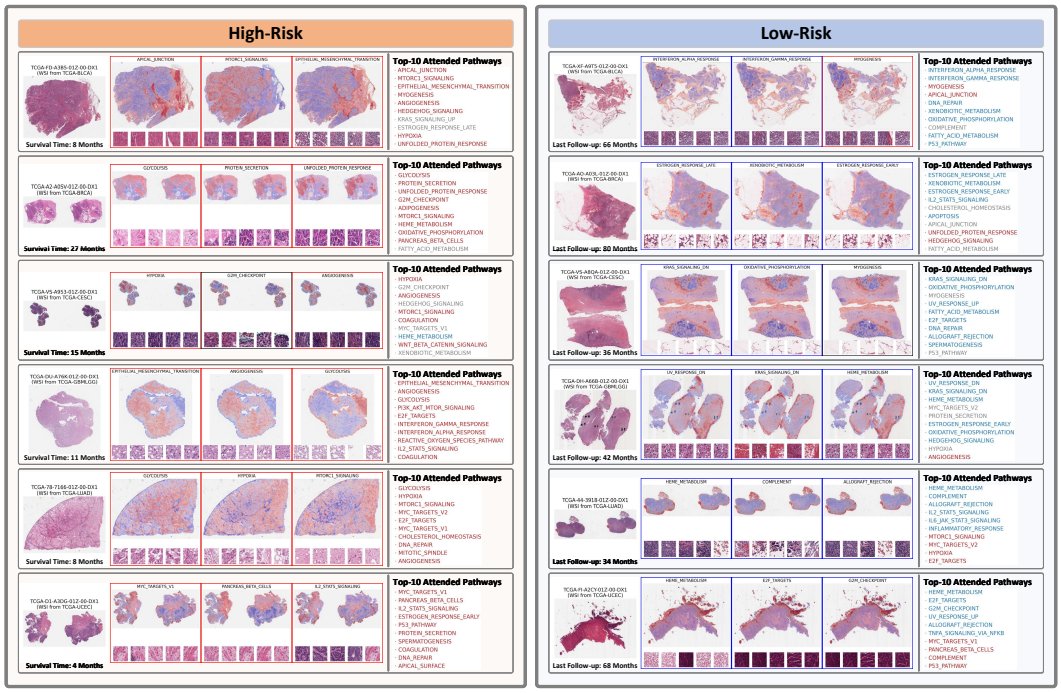

Figure 13: Attribution visualization across all cancer cohorts. The cross-attention weights are projected onto histopathological regions to illustrate pathway-specific contributions to patient risk prediction. Top-3 pathways with highest learned importance are attributed, with each attention map highlighting the top-5 patches with the highest attention scores. Red, blue, and gray indicate the top-15 promoting pathways, top-15 protective pathways, and remaining pathways, respectively.

The attribution results across all cancer cohorts are presented in Fig. 13. Taking the UCEC cohort as an example, the high-risk case exhibits distinct pathway-specific attribution patterns, where different pathways focus on complementary histopathological regions. The MYC Targets V1 pathway highlights regions with dense tumor cells and increased nuclear density, reflecting the association between MYC activation and tumor cell proliferation. The IL-2/STAT5 signaling pathway mainly attends to stromal regions enriched with immune-related components, suggesting that StructAlign captures microenvironment-associated signals related to patient risk.

Meanwhile, the PANCREAS BETA CELLS pathway emphasizes tumor regions with heterogeneous morphology, indicating that the model identifies diverse metabolic and morphological patterns associated with aggressive tumor characteristics. Together, these pathway-specific attributions demonstrate that StructAlign learns meaningful associations between pathway activities and spatial histopathological patterns, providing interpretable evidence for robust prediction.

Furthermore, we observe a consistent relationship between pathway attribution and the survival-associated pathway preferences analyzed in Sec. 5.5. For high-risk patients with shorter survival times, StructAlign tends to assign higher attention to pathways categorized as risk-promoting. In contrast, for low-risk patients with longer survival times, the model primarily relies on pathways with protective prognostic associations. This consistency indicates that StructAlign not only preserves pathway-level structural information but also effectively incorporates survival-associated pathway characteristics into patient-specific multimodal risk prediction.