PROTOTYPING PROGRESSIVE SURVIVAL-RISK INTO CROSS-MODAL REPRESENTATION LEARNING FOR MULTIMODAL CANCER SURVIVAL PREDICTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Multimodal cancer survival prediction integrates complementary information from pathological Whole Slide Images and other modalities, such as gene expression data, to estimate patient-specific survival risk. Existing methods mainly focus on individual-level cross-modal alignment, overlooking the shared risk structures among patients and the progressive nature of survival risk. To address these limitations, we propose Risk-Pro, a Progressive Survival-Risk Prototyping framework that introduces population-level risk structures into multimodal representation learning. Specifically, Risk-Pro constructs risk prototypes to capture shared prognostic patterns across patients with similar survival states and aligns patient representations with corresponding risk structures. Moreover, a Progressive Prototype Constraint is proposed to organize risk prototypes along a continuous survival trajectory, encouraging closer representations for adjacent risk states and greater separation for distant ones. Experiments on six TCGA cancer datasets demonstrate that Risk-Pro consistently outperforms state-of-the-art methods and effectively captures progressive survival-risk structures, as validated by risk trend evaluation, Kaplan-Meier survival analysis, and attribution analysis.

1 INTRODUCTION

Histopathological Whole Slide Images (WSIs) provide high-resolution insights into the tissue microenvironment, capturing critical morphological structures and cellular heterogeneity (Song et al., 2023; Lu et al., 2021b). While genomic data, such as gene expression profiles, offer complementary insights into molecular mechanisms like immune status (El Nahhas et al., 2024; Zeng et al., 2023). By integrating these modalities, **multimodal cancer survival prediction** aims to accurately estimate patient-specific survival risk by uncovering associations between pathological and molecular characteristics (Afonso et al., 2024; Győrffy, 2021). More recent advances have investigated cross-modal knowledge transfer by reconstructing or inferring latent gene representations from pathological data (Wang et al., 2025; Xu et al., 2025; Zhou et al., 2025), enabling models to exploit genomic knowledge acquired during training even when only pathology data are available at inference time.

However, existing approaches mainly emphasize patient-specific cross-modal alignment (Jaume et al., 2024; Zhang et al., 2024; Zhou et al., 2025), which captures modality consistency within individual cases but overlooks shared prognostic patterns across patients with similar survival profiles. This limitation hinders the discovery of population-level dependencies that characterize survival outcomes (Kent et al., 2018; 2020). Furthermore, survival risk is inherently **progressive rather than categorical**, evolving continuously from favorable to adverse states (e.g., safe $\rightarrow$ low $\rightarrow$ mid $\rightarrow$ high risk) (Ning et al., 2026). Ignoring the hierarchical dependencies among risk stages may lead to suboptimal risk representations that fail to reflect the continuous trajectory of disease progression.

To address these limitations, we propose **Risk-Pro**, a *Progressive Survival-Risk Prototyping* framework that introduces risk-aware population structures into cross-modal representation learning. Specifically, Risk-Pro constructs *risk-associated prototypes* from gene expression data to characterize shared biological patterns among patients with similar survival-risk profiles. These prototypes serve as population-level risk references to guide representation learning through a reconstruction objective, aligning patient representations with survival-risk states while capturing population-level

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.

risk structures beyond patient-level alignment. Furthermore, we introduce a *Progressive Prototype Constraint* that organizes risk prototypes according to the survival hierarchy through a contrastive objective, encouraging compact representations for adjacent risk states while enforcing greater separation between distant ones. By embedding such progressive survival-risk organization into the latent space, Risk-Pro captures clinically meaningful risk transitions and facilitates risk-aware multimodal cancer survival prediction.

To comprehensively evaluate Risk-Pro, we conduct extensive experiments across multiple cancer datasets. Risk-Pro consistently outperforms state-of-the-art competing methods, achieving an average improvement of 2.93% in C-index. The Jonckheere-Terpstra test and correlation analysis across different risk-level groups verify that the learned representations preserve the progressive ordering of survival risk. Furthermore, Kaplan-Meier analysis and attribution-based analysis demonstrate effective risk stratification and reveal specific prognostic features underlying different risk levels. In summary, our key contributions are as follows:

1. We introduce Risk-Pro, a *Progressive Survival-Risk Prototyping* framework that moves beyond existing patient-level modeling paradigms by constructing *risk-associated prototypes* as population-level guidance to capture shared prognostic patterns.
2. We propose a *Progressive Prototype Constraint* to organize these prototypes along the continuous risk structure, enabling the learned representation space to reflect meaningful transitions across different risk states.
3. Extensive experiments demonstrate that Risk-Pro consistently outperforms state-of-the-art methods, which is further supported by multiple complementary analyses.

2 Related Work

2.1 Multimodal Cancer Survival Prediction

Multimodal cancer survival prediction aims to quantify prognostic risk by jointly leveraging pathological slide images and gene expression data (Afonso et al., 2024; Győrffy, 2021).

Early studies primarily focus on cross-modality fusion paradigms (Ding et al., 2023; Xiong et al., 2024), where modality-specific encoders independently extract representations and combine them through various fusion mechanisms (Qu et al., 2025; Yan et al., 2025; Raza et al., 2025). Representative methods have investigated cross-attention mechanisms (Jaume et al., 2024; Zhou & Chen, 2023) and optimal transport-based alignment (Xu & Chen, 2023) to facilitate interactions between modalities. However, these approaches mainly emphasize distribution-level alignment, while overlooking the underlying cross-modal structural relationships.

More recently, due to the limited accessibility and high cost of genomic profiling in clinical practice (Payne et al., 2018), cross-modal knowledge transfer approaches have emerged (Ma et al., 2021; Lee et al., 2023; Wang et al., 2023). They leverage genomic information as auxiliary supervision during training to enhance pathological representation learning, enabling molecular knowledge to be leveraged with pathology-only inference (Xu et al., 2025; Zhou et al., 2025).

Despite these advances, existing methods mainly focus on modeling cross-modal consistency at the *individual level*, overlooking survival-related population structures shared among patients with similar outcomes (Kent et al., 2018; 2020). Explicit modeling of such population-wise structures remains underexplored, motivating the development of survival-aware representation learning frameworks.

2.2 Prototype Learning

Prototype learning represents samples with similar semantic characteristics through a compact set of prototypes (Snell et al., 2017). In this paradigm, each prototype serves as a semantic anchor for a specific category, and predictions are typically made by measuring the similarity between individual samples and these prototypes. By summarizing shared characteristics among samples, prototype learning enables models to capture underlying structures beyond individual-level representations and facilitates structured representation learning.

Given that the WSIs are extremely high-resolution, they are commonly pre-divided into a large number of patches to enable feasible computation. In survival prediction, prototype-based methods have been explored to reduce the computational burden by compressing numerous patch-level features into representative representations (Song et al., 2024a;b). Typically, these prototypes are obtained through clustering-based strategies to summarize patch features for multiple instance learning.

However, these prototype-based approaches mainly focus on the compression of feature dimensions, without modeling the explicit population-level survival structures among patients. In contrast, Risk-Pro constructs risk-associated prototypes based on survival characteristics, shifting the paradigm from *single-patient* modeling to *patient-prototype* representation learning for capturing risk-associated survival patterns among patients.

3 PRELIMINARIES

3.1 PROBLEM FORMULATION

In multimodal cancer survival prediction, a dataset containing $|\mathcal{D}|$ patients is represented as $\mathcal{D} = \{(X_i^{path}, X_i^{gene}, T_i, c_i)\}_{i=1}^{|\mathcal{D}|}$, where X_i^{path} and X_i^{gene} denote pathological and genomic features, respectively. T_i represents the observed follow-up time, and c_i is the censoring indicator ($c_i = 1$ for censored patients and $c_i = 0$ for deceased patients). For simplicity, we omit the patient index i in the following descriptions.

The goal of multimodal cancer survival prediction is to learn a prognostic function that estimates a patient-specific risk score from multimodal inputs:

$$\hat{r} = f(X^{path}, X^{gene}), \tag{1}$$

where $f(\cdot)$ denotes the prediction function. Model performance is commonly evaluated by the **C-index** (Harrell Jr et al., 1996) that measures the consistency between predicted risk scores and observed survival outcomes. Detailed computation of C-index can be found in Appendix A.

3.2 DATA PRE-PROCESSING

As illustrated in the upper panel of Fig. 1, we perform the following pre-processing steps prior to model training.

Pathology features. To enable feasible computation, each WSI is divided into multiple patches. Each patch is encoded into a vector using a *fixed encoder*, we employ the pre-trained pathological foundation model CTransPath (Wang et al., 2021; 2022) by default. This process transforms each WSI into a pathological feature matrix $X^{path} \in \mathbb{R}^{N \times d}$, where N denotes the number of patches and d is the patch-level feature dimension.

Genomic features. Consistent with established bioinformatic protocols (Chen et al., 2021), genes are grouped into M biologically informed bags, with the m-th gene bag represented as $\tilde{X}_m^{gene}$. To unify feature dimensions across bags, a trainable gene encoder Ω_m (implemented as a linear layer by default) is applied to each bag, mapping its features to the shared dimension d. The resulting genomic feature matrix is denoted as $X^{gene} \in \mathbb{R}^{M \times d}$.

Risk-level partition and prototype construction. To explicitly capture risk patterns, we partition patients into K risk-level groups (e.g., safe, low, mid-low, mid-high, high) based on the quantiles of observed times and censoring states. [1]

For each gene bag m, we compute **gene prototypes** by averaging genomic features of patients within the same risk-level group, yielding $\tilde{P}_m^k$, where k indexes the risk level. The encoded representation of each prototype is denoted as $P_m^k = \Omega_m(\tilde{P}_m^k)$. To avoid data leakage, these prototypes are aggregated exclusively from the training set $\mathcal{D}_{train}$.

[1] Detailed definition of risk-level partition and sensitivity analysis of K are in Appendix A.

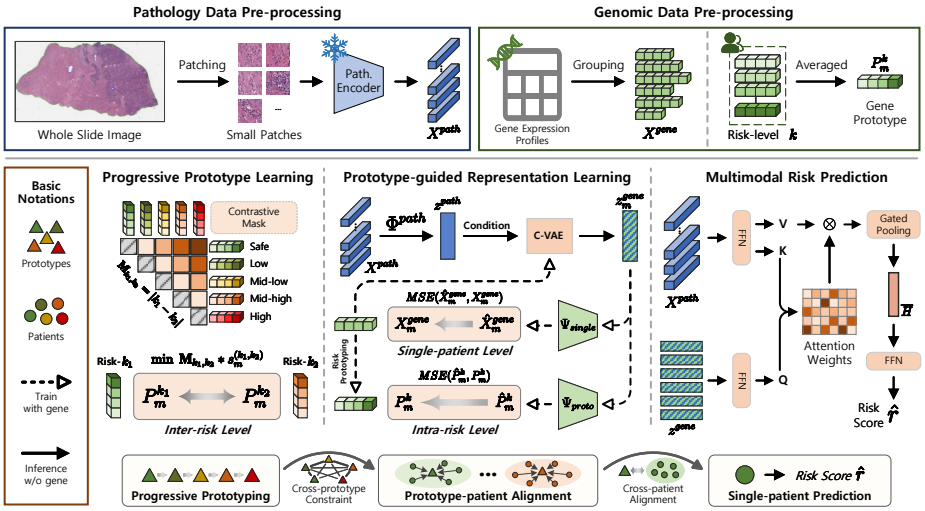

Figure 1: Architectural overview of the proposed Risk-Pro framework.

4 METHODOLOGY

In this section, we introduce the **Risk-Pro** framework. As shown in Fig. 1, it consists of three main components: (1) Prototype-guided Representation Learning (Sec. 4.1), (2) Progressive Prototype Learning (Sec. 4.2), and (3) Multimodal Risk Prediction (Sec. 4.3).

4.1 PROTOTYPE-GUIDED REPRESENTATION LEARNING

To enable pathology-only inference while retaining genomic knowledge, we learn genomic representations through pathology-genomics cross-modal learning during training. Given pathological features X^{path} consisting of multiple patch-level embeddings, a transformer-based encoder Φ^{path} is first employed to aggregates these patch-level features into a slide-level representation $z^{path} \in \mathbb{R}^d$:

$$z^{path} = \Phi^{path}(X^{path}). \tag{2}$$

Following LD-CVAE (Zhou et al., 2025), we employ a conditional variational autoencoder (C-VAE) architecture to model the cross-modality dependency between pathology and genomics. Specifically, for each gene bag m, the latent representation z_m^{gene} is learned from the genomic feature X_m^{gene}, conditioned on the pathology representation z^{path}:

$$z_m^{gene} = \text{C-VAE}(X_m^{gene}; z^{path}). \tag{3}$$

During inference, when gene expression data are unavailable, the genomic input X_m^{gene} is removed, and the learned cross-modal dependency transfers genomic knowledge from training to inference, enabling the latent genomic representation to be inferred based on the condition z^{path}.

To introduce survival-aware guidance, Risk-Pro incorporates risk-associated prototypes as population-level structural supervision. The genomic representation z_m^{gene} is optimized to preserve both *patient-specific characteristics* and *intra-risk patterns* through a dual-level reconstruction.

From the **patient-specific** perspective, the genomic feature of each patient is reconstructed from z_m^{gene} via a patient-specific decoder $\Psi patient$:

$$\hat{X}_m^{gene} = \Psi_{patient}(z_m^{gene}). \tag{4}$$

Meanwhile, to capture **intra-risk** patterns among patients within the same risk-level, the corresponding risk prototype is reconstructed via a prototype-specific decoder Ψ_{proto}:

$$\hat{P}_m^k = \Psi_{proto}(z_m^{gene}), \tag{5}$$

where k denotes the patient's corresponding risk level.

By jointly reconstructing individual genomic features and risk prototypes, Risk-Pro learns genomic representations that retain patient-specific characteristics while encoding population-level structures. The C-VAE framework is optimized by minimizing the negative Evidence Lower Bound (ELBO):

$$\mathcal{L}_{ELBO} = \sum_{m=1}^{M} \Big(\mathbb{E}_{q(z_m^{gene}|X_m^{gene}, z^{path})} \Big[MSE(\hat{X}_m^{gene}, X_m^{gene}) + MSE(\hat{P}_m^k, P_m^k) \Big] + KL(q(z_m^{gene}|X_m^{gene}, z^{path}) \,||\, p(z_m^{gene})) \Big), \tag{6}$$

where $MSE(\cdot)$ denotes the mean squared error, $p(\cdot)$ and $q(\cdot)$ are the prior and posterior distributions, respectively. The first two terms enforce accurate reconstruction of the patient-specific genomic feature X_m^{gene} and the prototypical genomic features P_m^k, while the last term regularizes the latent space of z_m^{gene} through the Kullback-Leibler (KL) divergence.

4.2 Progressive Prototype Learning

Survival risk exhibits an inherently *progressive structure*, where risk evolve in an ordered manner from favorable to adverse outcomes (Ning et al., 2026). To explicitly capture the hierarchical relationships, we introduce a **Progressive Prototype Constraint** that regularizes the organization of risk prototypes according to their ordinal relationships.

To encode the relative distance between risk levels, we define an *ordinal distance matrix* $\mathbb{M} \in \mathbb{R}^{K \times K}$. For the m-th gene bag, we measure the similarity between two risk prototypes corresponding to risk levels k_1 and k_2 using cosine similarity:

$$\mathbb{M}_{k_1,k_2} = |k_1 - k_2|, \quad s_m^{(k_1,k_2)} = \cos(P_m^{k_1}, P_m^{k_2}). \tag{7}$$

To obtain risk-discriminative representations, we impose an ordinal constraint on prototype similarities, encouraging prototypes with closer risk levels to remain more similar while enforcing larger separation between prototypes with greater risk distances, the objective is formulated as:

$$\mathcal{L}_{Prog} = \sum_{m=1}^{M} \sum_{k_1=1}^{K} \sum_{k_2=1}^{K} \mathbb{I}(k_1 > k_2)(\mathbb{M}_{k_1,k_2} * s_m^{(k_1,k_2)}), \tag{8}$$

where $\mathbb{I}(\cdot)$ is an indicator function ensuring each pair of risk levels is considered only once. By imposing the progressive prototype constraint, Risk-Pro organizes risk prototypes according to their ordinal relationships, enabling the learned representations to capture continuous structures.

4.3 Multimodal Risk Prediction

We employ cross-attention (Vaswani et al., 2017) to fuse the two modalities, where z^{gene} serves as the query and X^{path} as the key and value:

$$H = \text{Attention}(Q = z^{gene}; K = V = X^{path}). \tag{9}$$

This design enables the model to aggregate patch-level pathological information from the genomic perspective, producing fused representations $H \in \mathbb{R}^{M \times d}$. To obtain a global representation, we apply gated pooling over the fused features $\overline{H} \in \mathbb{R}^d$:

$$a_m = \frac{\exp(\text{FFN}(H_m))}{\sum_{m=1}^{M} \exp(\text{FFN}(H_m))}, \quad \overline{H} = \sum_{m=1}^{M} a_m * H_m, \tag{10}$$

where H_m denotes the fused feature corresponding to the m-th gene perspective, and a_m is its normalized attention weight computed via a feed-forward network (FFN) followed by softmax. The aggregated vector $\overline{H} \in \mathbb{R}^d$ summarizes the cross-modal information.

Finally, the patient-specific risk score $\hat{r} \in (0, 1)$ is predicted from the aggregated representation $\overline{H}$ using another FFN followed by a sigmoid activation $\sigma(\cdot)$:

$$\hat{r} = \sigma(\text{FFN}(\overline{H})), \quad \mathcal{L}_{NLL} = NLL(\hat{r}), \tag{11}$$

where $\mathcal{L}_{NLL}$ represents the negative log-likelihood loss for optimization. Detailed computation for $\mathcal{L}_{NLL}$ is provided in Appendix A. The overall optimization objective is defined as:

$$\mathcal{L} = \mathcal{L}_{NLL} + \lambda_1 * \mathcal{L}_{ELBO} + \lambda_2 * \mathcal{L}_{Prog}. \tag{12}$$

Here, λ_1 and λ_2 are hyper-parameters. $\mathcal{L}_{NLL}$ optimizes survival prediction, $\mathcal{L}_{ELBO}$ learns risk-associated representations at both the **single-patient** and **intra-risk** levels, and $\mathcal{L}_{Prog}$ enforces progressive prototype structures at the **inter-risk** level.

5 Experiments

5.1 Benchmark Settings

We conduct experiments on six cancer datasets from The Cancer Genome Atlas (TCGA) (Weinstein et al., 2013), including BLCA, BRCA, CESC, GBMLGG, LUAD, and UCEC. Each dataset contains WSIs, matched bulk gene expression profiles, and clinical metadata with overall survival. We compare Risk-Pro with the following baselines:

1) Unimodal settings. For the **gene-only** setting, we include MLP Haykin (1994), SNN Klambauer et al. (2017), and SNNTrans Klambauer et al. (2017); Shao et al. (2021). For **pathology-only** prediction, we include ABMIL (Ilse et al., 2018), CLAM (Lu et al., 2021a) (SB and MB variants), and TransMIL (Shao et al., 2021).

2) Multimodal setting. We compare with 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).

Experiments are conducted via 5-fold cross-validation. The loss weights in Eq. (12) are set to $\lambda_1 = 0.5$ and $\lambda_2 = 0.5$, and the number of risk levels is set to 5. Detailed implementations and hyperparameter analyses are in Appendix A.

Table 1: C-index comparison across six cancer datasets, with N denoting the number of patients. g. and p. indicate genomic and pathological modalities, respectively. Values are reported as mean ± standard deviation (std). **Bold** values indicate the best and underline values indicate the runner-up.

Modality	Method	BLCA ($N = 373$)	BRCA ($N = 957$)	CESC ($N = 241$)	GBMLGG ($N = 571$)	LUAD ($N = 452$)	UCEC ($N = 480$)
g.	MLP	0.606±0.039	0.645±0.038	0.665±0.062	0.809±0.015	0.601±0.060	0.556±0.113
	SNN	0.600±0.024	0.620±0.056	0.657±0.073	0.804±0.057	0.596±0.058	0.546±0.037
	SNNTrans	0.619±0.030	0.633±0.036	0.653±0.068	0.805±0.028	0.618±0.050	0.664±0.068
p.	ABMIL	0.538±0.043	0.617±0.049	0.610±0.044	0.766±0.043	0.576±0.029	0.585±0.033
	CLAM-SB	0.576±0.052	0.609±0.026	0.657±0.061	0.793±0.036	0.603±0.028	0.675±0.034
	CLAM-MB	0.598±0.064	0.621±0.039	0.677±0.089	0.794±0.033	0.620±0.022	0.661±0.053
	TransMIL	0.574±0.051	0.637±0.024	0.708±0.092	0.795±0.042	0.607±0.028	0.640±0.052
p.+ g.	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
	Risk-Pro	**0.658±0.069**	**0.683±0.021**	**0.753±0.059**	**0.848±0.017**	**0.658±0.038**	**0.712±0.034**

5.2 Main Results

To evaluate the effectiveness and generalizability of Risk-Pro, we conduct experiments on six cancer datasets and compare its performance with a range of baseline methods. Table 1 reports the C-index results based on the 5-fold cross-validation, showing that **Risk-Pro consistently outperforms all baseline methods**, demonstrating its robustness and generalizability across diverse cancer types. The improvement can be attributed to two key designs that capture risk structures from both the intra-risk and inter-risk perspectives:

1) Risk-aware prototype-guided learning. Risk-Pro leverages risk-associated prototypes as structured guidance to learn risk-aware representations at both the **single-patient** and **intra-risk** levels, capturing patient-specific characteristics and shared prognostic patterns.

2) Progressive prototype learning. By enforcing structured relationships among risk prototypes, Risk-Pro captures the continuous evolution of survival risk and learns more risk-discriminative representations at the **inter-risk** level.

Together, these two designs enable Risk-Pro to model survival risk from complementary perspectives, capturing both shared risk patterns within patient groups and progressive relationships across different risk states.

Table 2: Ablation evaluation of Risk-Pro compared with alternative variants. "w/o gene" indicates the gene representations are reconstructed in the scenario of unavailable genomic data.

Variant	w/o gene	BLCA	BRCA	CESC	GBMLGG	LUAD	UCEC
Risk-Pro		**0.658±0.069**	**0.683±0.021**	**0.753±0.059**	**0.848±0.017**	**0.658±0.038**	**0.712±0.034**
	✓	0.656±0.055	0.655±0.014	0.736±0.101	0.827±0.028	0.654±0.032	0.707±0.021
Intra-risk		0.646±0.071	0.674±0.034	0.723±0.104	0.840±0.017	0.636±0.022	0.706±0.027
	✓	0.651±0.058	0.650±0.036	0.703±0.096	0.821±0.035	0.638±0.023	0.699±0.013
Single-patient		0.641±0.061	0.637±0.022	0.715±0.021	0.812±0.034	0.624±0.029	0.707±0.028
	✓	0.637±0.058	0.625±0.040	0.693±0.019	0.798±0.034	0.620±0.036	0.702±0.039

5.3 Ablation Study

To evaluate the contribution of each component in Risk-Pro, we compare it with the two variants:

- **Intra-risk**: removes the Progressive Prototype Constraint while retaining gene reconstruction at both the *single-patient* and *intra-risk* levels, eliminating *inter-risk* relationships among risk states.
- **Single-patient**: performs gene reconstruction only at the *single-patient* level, without incorporating risk-associated prototypes for cross-patient modeling.

Furthermore, to simulate gene-missing inference scenarios, we reconstruct the gene representations z^{gene} only based on the condition z^{path}. As shown in Table 2, Risk-Pro consistently outperforms all ablated variants, validating the effectiveness of its key components. Even without genomic inputs, Risk-Pro maintains superior performance, demonstrating its robustness in pathology-only inference.

Removing the Progressive Prototype Constraint (i.e., the **Intra-risk** variant) degrades performance, highlighting the importance of modeling progressive relationships among risk states. When the risk-associated reconstruction mechanism is further removed (i.e., the **Single-patient** variant), the performance declines even more significantly, indicating that modeling cross-patient risk relationships enables Risk-Pro to effectively capture the shared molecular patterns.

5.4 Validation of Progressive Risk Structure

To further validate the effectiveness of risk-aware modeling in Risk-Pro, we analyze both the risk scores and latent gene representations with respect to the risk-level groups.

5.4.1 Jonckheere-Terpstra Test on Predicted Risk Scores

To evaluate whether the risk scores predicted by Risk-Pro exhibit a monotonic ordering across increasing risk levels, we perform the Jonckheere-Terpstra (JT) (Jonckheere, 1954; Terpstra, 1952) trend test. Specifically, we collect the risk scores for patients according to their ground-truth risk levels and apply the JT test to assess whether these scores exhibit a monotonic ascending trajectory consistent with the clinical progression.

As illustrated in Fig. 2, the predicted risk scores generally exhibit an increasing trend as the risk levels advance from safe to high risk, indicating that Risk-Pro successfully learns representations that align with the underlying progression of patient risk levels. These results further support the effectiveness of progressive risk-associated modeling, demonstrating that Risk-Pro captures the evolutionary nature of survival risk in a statistically robust and consistent manner.

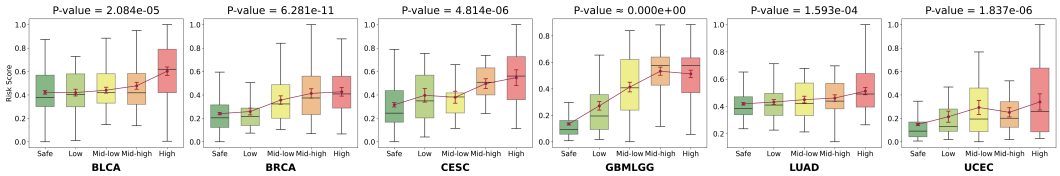

Figure 2: The Jonckheere-Terpstra test results on risk-level groups.

5.4.2 Correlation Analysis of Latent Gene Representations

To examine the organization of the learned representation space, we analyze the correlations among genomic representations across different survival-risk levels. Specifically, the latent representations z_m^{gene} of all gene bags are globally averaged to obtain a patient-level representation for analysis.

From the heatmaps in Fig. 3, a trend is observed: representations of adjacent risk levels exhibit higher correlations, whereas more distant risk levels are less correlated. This pattern is particularly evident in the BRCA, GBMLGG, and UCEC datasets. These observations indicate that explicitly modeling risk-associated relationships is crucial for capturing meaningful similarities among patient groups and preserving the hierarchical structure of survival.

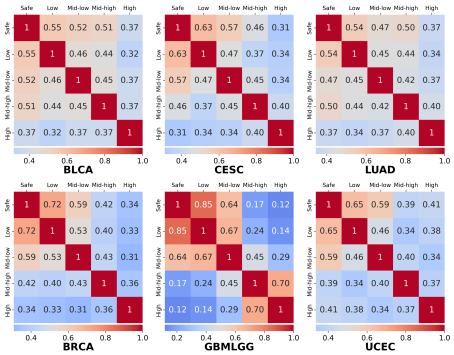

Figure 3: Correlations analysis of the latent gene representations.

5.5 Kaplan-Meier Overall Survival Analysis

In addition to C-index evaluations, we further stratify patients equal-sized high- and low-risk groups based on the risk scores predicted by Risk-Pro and assess their differences using Kaplan-Meier (KM) curves (Kaplan & Meier, 1958) under both complete-modality and gene-missing scenarios.

The results and Log-rank test (Mantel et al., 1966) p-values are shown in Fig. 4. Risk-Pro consistently achieves clear separation between high- and low-risk groups in both settings, demonstrating its ability to capture clinically meaningful survival patterns. These results further validate the effectiveness of modeling risk-associated structures, enabling robust and interpretable survival prediction.

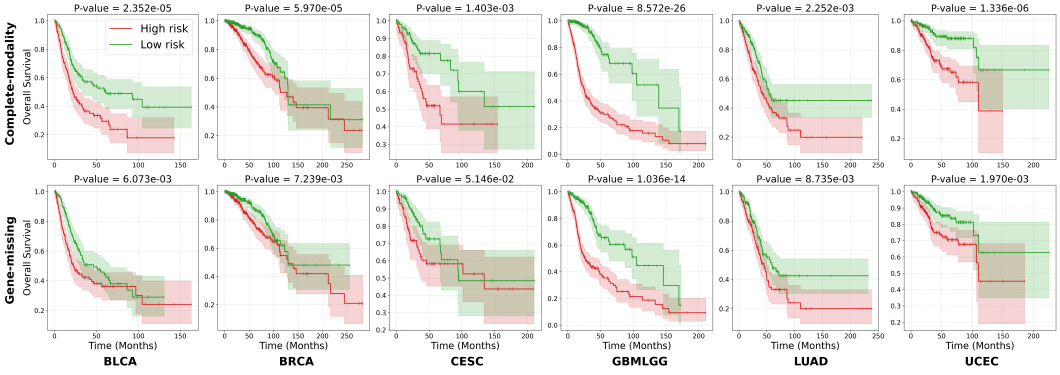

Figure 4: Kaplan-Meier overall survival analysis of the predicted high-risk (red) and low-risk (green) groups under both the complete-modality (top) and gene-missing (bottom) scenarios. Confidence intervals are shown as shaded regions.

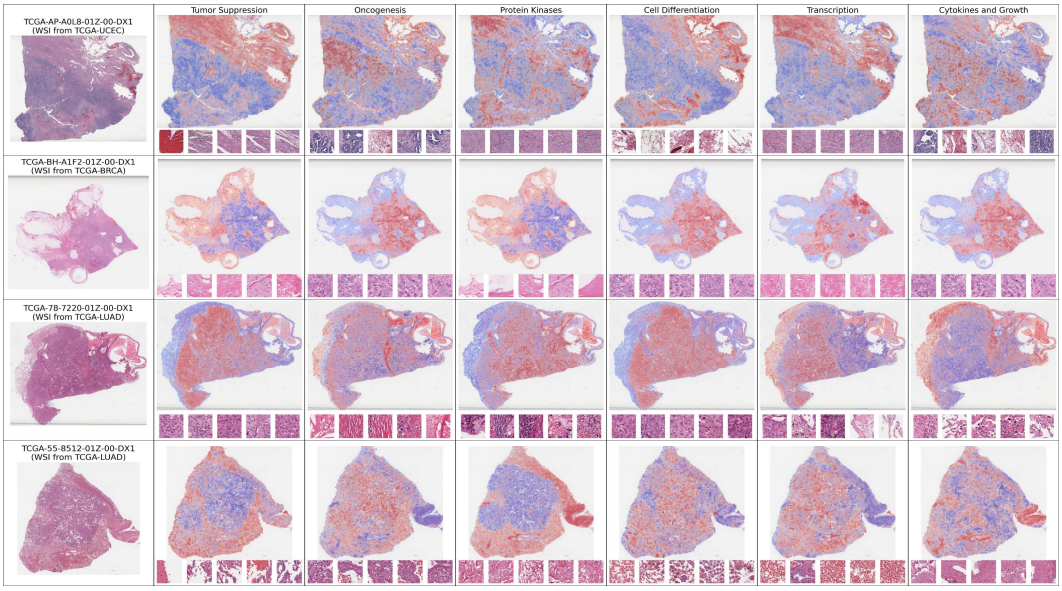

Figure 5: Visualization of cross-attention weights in Risk-Pro. Each column corresponds to a gene bag. The images shown below each heatmap represent the top-5 patches with the highest attention weights corresponding to the gene bag.

5.6 Attribution Visualization

To further investigate how Risk-Pro captures associations between pathological structures and genomic signatures, we visualize the spatial distribution of cross-attention weights in four representative high-risk cases defined in Sec. 5.5. All interpretations were reviewed by expert pathologists and confirmed to be consistent with known pathological morphology. As shown in Fig. 5, the attention maps exhibit distinct spatial patterns across different functional gene categories. [2]

For oncogenic pathways such as *Oncogenesis* and *Cytokines and Growth*, high-attention regions are concentrated in tumor parenchyma and proliferative cancer nests, whereas tumor suppressor pathways exhibit broader distributions across cancer regions. In LUAD cases, oncogenic pathway attention focuses on the tumor core, while tumor suppressor pathway attention shows a more diffuse pattern. Overall, these results demonstrate that Risk-Pro captures interpretable and biologically meaningful cross-modal associations.

6 Conclusion

In this work, we propose **Risk-Pro**, a *Progressive Survival-risk Prototyping* framework for multimodal cancer survival prediction. Risk-Pro introduces risk-associated prototypes as globally structured guidance to capture shared prognostic patterns among patients with similar survival risks. Through prototype-guided representation learning, Risk-Pro jointly preserves patient-specific characteristics and risk-level structures, enabling risk-aware cross-modal alignment. Moreover, the Progressive Prototype Constraint organizes risk prototypes according to the ordinal relationships of survival risk, enabling progressive risk modeling across levels.

Extensive experiments on TCGA datasets demonstrate that Risk-Pro consistently outperforms state-of-the-art methods and effectively learns progressive survival-risk representations. Comprehensive analyses, including the Jonckheere-Terpstra test, correlation analysis, Kaplan-Meier survival analysis, and cross-modal attribution analysis, further validate its ability to preserve hierarchical survival-risk organization and interpret cross-modal prognostic patterns.

[2] More analyses of low-risk cases are provided in Appendix B.

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.

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.

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.

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.

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.

Simon Haykin. *Neural networks: a comprehensive foundation*. Prentice hall PTR, 1994.

Maximilian Ilse, Jakub Tomczak, and Max Welling. Attention-based deep multiple instance learning. In *International conference on machine learning*, pp. 2127–2136. PMLR, 2018.

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.

Aimable Robert Jonckheere. A distribution-free k-sample test against ordered alternatives. *Biometrika*, 41(1/2):133–145, 1954.

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

David M Kent, Ewout Steyerberg, and David Van Klaveren. Personalized evidence based medicine: predictive approaches to heterogeneous treatment effects. *Bmj*, 363, 2018.

David M Kent, Jessica K Paulus, David Van Klaveren, Ralph D'Agostino, Steve Goodman, Rodney Hayward, John PA Ioannidis, Bray Patrick-Lake, Sally Morton, Michael Pencina, et al. The predictive approaches to treatment effect heterogeneity (path) statement. *Annals of internal medicine*, 172(1):35–45, 2020.

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

Günter Klambauer, Thomas Unterthiner, Andreas Mayr, and Sepp Hochreiter. Self-normalizing neural networks. *Advances in neural information processing systems*, 30, 2017.

Yi-Lun Lee, Yi-Hsuan Tsai, Wei-Chen Chiu, and Chen-Yu Lee. Multimodal prompting with missing modalities for visual recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14943–14952, 2023.

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, 2021a.

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, 2021b.

Mengmeng Ma, Jian Ren, Long Zhao, Sergey Tulyakov, Cathy Wu, and Xi Peng. Smil: Multimodal learning with severely missing modality. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pp. 2302–2310, 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.

Catherine Ning, Dimitris Bertsimas, Per Eystein Lønning, Federico N Auecio, Richard Burkhart, Felix Balzer, Stefan Buettner, Hideo Baba, Itaru Endo, Georgios Stasinos, et al. Improving survival models in health care by balancing imbalanced cohorts: A novel approach. *JCO clinical cancer informatics*, 10:e2500190, 2026.

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.

Katherine Payne, Sean P Gavan, Stuart J Wright, and Alexander J Thompson. Cost-effectiveness analyses of genetic and genomic diagnostic tests. *Nature Reviews Genetics*, 19(4):235–246, 2018.

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.

Manahil Raza, Ayesha Azam, Talha Qaiser, and Nasir Rajpoot. Ps3: A multimodal transformer integrating pathology reports with histology images and biological pathways for cancer survival prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 22175–22186, 2025.

Zhuchen Shao, Hao Bian, Yang Chen, Yifeng Wang, Jian Zhang, Xiangyang Ji, et al. Transmil: Transformer based correlated multiple instance learning for whole slide image classification. *Advances in neural information processing systems*, 34:2136–2147, 2021.

Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. *Advances in neural information processing systems*, 30, 2017.

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.

Andrew H Song, Richard J Chen, Tong Ding, Drew FK Williamson, Guillaume Jaume, and Faisal Mahmood. Morphological prototyping for unsupervised slide representation learning in computational pathology. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11566–11578, 2024a.

Andrew H. Song, Richard J. Chen, Guillaume Jaume, Anurag Jayant Vaidya, Alexander Baras, and Faisal Mahmood. Multimodal prototyping for cancer survival prediction. In *Forty-first International Conference on Machine Learning*, 2024b. URL https://openreview.net/forum?id=3MfvxH3Gia.

Teunis J Terpstra. The asymptotic normality and consistency of kendall's test against trend, when ties are present in one ranking. *Indagationes Mathematicae*, 14(3):327–333, 1952.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *NeurIPS*, 2017.

Hu Wang, Yuanhong Chen, Congbo Ma, Jodie Avery, Louise Hull, and Gustavo Carneiro. Multimodal learning with missing modality via shared-specific feature modelling. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 15878–15887, 2023.

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.

Zhikang Wang, Yumeng Zhang, Yingxue Xu, Seiya Imoto, Hao Chen, and Jiangning Song. Histogenomic knowledge association for cancer prognosis from histopathology whole slide images. *IEEE Transactions on Medical Imaging*, 44(5):2170–2181, 2025.

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.

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.

Yingxue Xu, Fengtao Zhou, Chenyu Zhao, Yihui Wang, Can Yang, and Hao Chen. Distilled prompt learning for incomplete multimodal survival prediction. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 5102–5111, 2025.

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.

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 Implementation Details

A.1 Cancer Datasets

The TCGA cancer datasets used in our study include:

- Bladder Urothelial Carcinoma (BLCA)
- Breast Invasive Carcinoma (BRCA)
- Cervical Squamous Cell Carcinoma (CESC)
- Glioblastoma and Lower Grade Glioma (GBMLGG)
- Lung Adenocarcinoma (LUAD)
- Uterine Corpus Endometrial Carcinoma (UCEC)

A.2 Genomic Data Implementation

Following prior studies (Jaume et al., 2024; Zhou et al., 2025), we partition the gene expression data into six biologically-informed gene bags:

- Tumor Suppression
- Oncogenesis
- Protein Kinases
- Cell Differentiation
- Transcription
- Cytokines and Growth

The genomic dataset comprises mutation (MUT) profiles, copy number variations (CNV), and RNA sequencing data. These data are publicly accessible via the cBioPortal database.

A.3 Evaluation Metric

The concordance index (C-index) quantifies the agreement between predicted risk scores and the observed ordering of patient survival times.

Unlike metrics that assess absolute prediction accuracy at the individual level, the C-index evaluates the consistency of relative risk ranking across patients. Formally, for a pair of patients (i, j), the pair is considered **comparable** if:

1. At least one patient is uncensored; and
2. The uncensored patient has a shorter observed survival time than the other.

A comparable pair is deemed **concordant** if the patient with the higher predicted risk score indeed has a shorter survival time, i.e.,

$$\hat{r}_i > \hat{r}_j \quad \text{if} \quad T_i < T_j. \tag{13}$$

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

$$\text{C-index} = \frac{\text{\# of concordant pairs}}{\text{\# of comparable pairs}}. \tag{14}$$

A.4 Experimental Settings

Due to the large scale of WSIs, feature extraction involves multi-stage preprocessing pipelines, including tissue patching and feature encoding). To ensure absolute fairness, all baselines and Risk-Pro were rigorously evaluated **under identical experimental conditions**, using the exact same locally pre-processed features and cross-validation splits.

All models are implemented in PyTorch (Paszke et al., 2017) and trained on a single NVIDIA A800 GPU. Due to the large size of pathology features, a batch size of 1 is employed, with gradient accumulation over 16 iterations to stabilize optimization.

The Adam optimizer (Kingma & Ba, 2014) is used with an initial learning rate of 2×10^{-4} and a weight decay of 1×10^{-5}. For pathology preprocessing, whole slide images (WSIs) are divided into 512×512 patches using the code from CLAM (Lu et al., 2021a) and CTransPath (Wang et al., 2022). Primary implementation scripts are provided in the Supplemental Materials.

For time-aware modeling and stable optimization, we follow prior studies (Jaume et al., 2024; Zhang et al., 2024; Zhou et al., 2025) by dividing the overall observation period of the dataset into four quartile-based time intervals, denoted as $t_1, \ldots, t_4$. Patient-specific risk is then predicted independently within each time interval, yielding four risk scores $(\hat{r}^{(t_1)}, \ldots, \hat{r}^{(t_4)})$ for each patient.

These predicted scores represent the relative likelihood that the patient has experienced the event within the corresponding time interval. During inference, these interval-wise risk scores are averaged into a unified risk score $\hat{r} = \frac{1}{4}\sum_{j=1}^{4} \hat{r}^{(t_j)}$.

A.5 Negative Log-likelihood Loss

For risk prediction, we adopt the negative log-likelihood (NLL) (Zadeh & Schmid, 2020) loss commonly used in survival analysis. We first define a survival time label $\tau \in \{1, 2, 3, 4\}$ for each patient, indicating the time interval in which the observed survival time falls:

$$\tau = j, \quad \text{if } t_{j-1} < T \leq t_j, \tag{15}$$

where c and T are the censoring indicator and observed survival time, respectively (as described in Sec. 3.1). The survival probability up to the j-th interval is computed as:

$$\hat{s}^{(t_j)} = \prod_{k=1}^{j} (1 - \hat{r}^{(t_k)}). \tag{16}$$

The patient-specific NLL loss is then formulated as:

$$\mathcal{L}_{NLL} = -c * \log(\hat{s}^{(t_\tau)}) - (1 - c) * (\log(\hat{s}^{(t_{\tau-1})}) + \log(\hat{r}^{(t_\tau)})). \tag{17}$$

A.6 Module Implementation

Pathology encoder The pathology latent representation z^{path} is obtained using a transformer-based encoder Φ^{path}:

$$z^{path} = \Phi^{path}(X^{path}) = \text{Transformer}([\texttt{<cls>}, X^{path}])[0], \tag{18}$$

where the output corresponding to the `<cls>` token serves as the global pathology embedding.

Conditional variational autoencoder For each gene bag m, a **C-VAE** is employed to learn the patient-specific genomic latent representation conditioned on the pathology embedding z^{path}.

The encoder produces the mean μ_m and log-variance $\log \sigma_m^2$ of a multivariate Gaussian distribution:

$$[\mu_m, \log \sigma_m^2] = \text{FFN}_{enc}([X_m^{gene}; z^{path}]), \tag{19}$$

where FFN_{enc} denotes a feed-forward network encoder and $[\,;\,]$ indicates concatenation.

The latent representation z_m^{gene} is obtained via the standard reparameterization operation:

$$z_m^{gene} = \mu_m + \sigma_m \odot \epsilon, \quad \epsilon \sim \mathcal{N}(0, I_d), \tag{20}$$

where $\odot$ denotes element-wise multiplication, and ϵ is sampled from the standard normal distribution.

Reconstruction decoder $\Psi_{patient}$ and Ψ_{proto} reconstruct the patient-specific gene features and risk-level prototypes, respectively, from z_m^{gene}. They are implemented as FFNs with identical architecture but independent parameters:

$$\hat{X}_m^{gene} = \Psi_{patient}(z_m^{gene}), \quad \hat{P}_m^k = \Psi_{proto}(z_m^{gene}), \tag{21}$$

where $\hat{X}_m^{gene}$ and $\hat{P}_m^k$ denotes the reconstructed patient-specific feature and the corresponding prototypical feature.

A.7 Risk-level Partition

Definition of Risk Levels The risk-level partition is used *only during training* for prototype learning and is not required during inference. To avoid information leakage, prototypes are constructed independently within each cross-validation fold using only the corresponding training set.

Specifically, the training survival time range is divided into $K-1$ intervals. Uncensored patients are assigned according to observed survival times, while censored patients with follow-up times beyond half of the overall time are assigned to the 'safe' group.

Sensitivity Analysis of Risk-level Partition We further tested different values of K. As shown in the table below, the performance of Risk-Pro generally exhibits an increase-then-decrease trend as K varies. Importantly, **Risk-Pro consistently outperforms LD-CVAE across all evaluated values of** K, suggesting the performance gain is not tied to a specific partition granularity, $K = 5$ achieves the best overall performance and is therefore adopted as the default configuration.

Table 3: Sensitivity analysis of risk-level partition.

K	BLCA	BRCA	CESC	GBMLGG	LUAD	UCEC
3	0.651±0.063	0.664±0.061	0.731±0.072	0.845±0.031	0.643±0.036	0.712±0.024
4	0.655±0.076	0.663±0.031	0.748±0.076	0.846±0.021	0.648±0.031	**0.716±0.015**
5	**0.658±0.069**	**0.683±0.021**	0.753±0.059	**0.848±0.017**	**0.658±0.038**	0.712±0.034
6	0.648±0.063	0.678±0.021	**0.758±0.049**	0.844±0.022	0.653±0.035	0.708±0.026
7	0.643±0.068	0.658±0.016	0.740±0.070	0.846±0.026	0.640±0.031	0.707±0.017
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

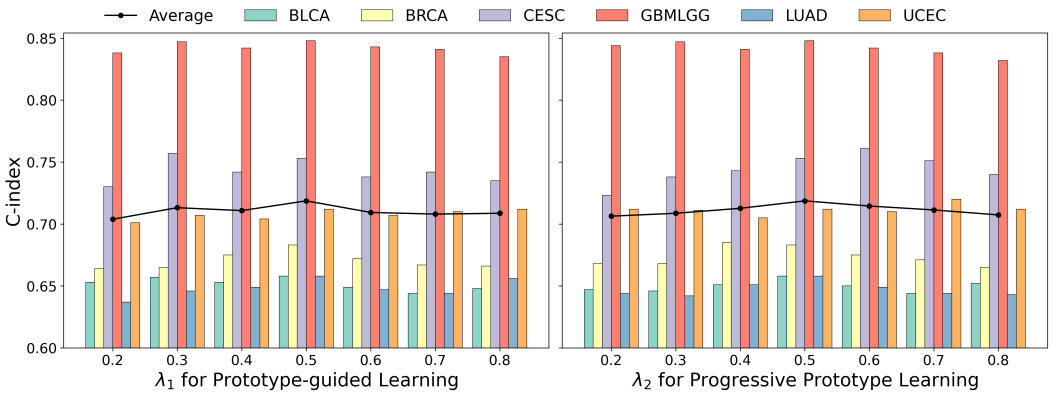

Figure 6: Sensitivity analysis of the loss weights λ_1 and λ_2 for the Prototype-guided Representation Learning objective $\mathcal{L}_{\text{ELBO}}$ and the Progressive Prototype Constraint $\mathcal{L}_{\text{Prog}}$, respectively.

A.8 Sensitivity Analysis of Loss Weights

We investigate the sensitivity of the loss weights λ_1 and λ_2, which control the contributions of the Prototype-guided Representation Learning objective $\mathcal{L}_{\text{ELBO}}$ and the Progressive Prototype Constraint $\mathcal{L}_{\text{Prog}}$, respectively. Both weights are varied from 0.2 to 0.8, with one weight fixed at 0.5 when evaluating the sensitivity of the other.

As shown in Fig. 6, although individual cancer datasets exhibit slightly different trends, the average performance consistently follows an increasing-then-decreasing pattern as the loss weights vary. This indicates that excessively small weights may provide insufficient guidance from the corresponding learning objective, whereas overly large weights may introduce an imbalance among different objectives. The best overall performance is achieved with $\lambda_1 = 0.5$ and $\lambda_2 = 0.5$, which are adopted as the default settings.

B Attribution Analysis

We further analyze attention distributions in low-risk patients. As shown in Fig. 7, Risk-Pro identifies distinct high-attention regions within oncogenic signaling pathways, while tumor suppressor pathways exhibit broader spatial activation across the entire cancer nest architecture.

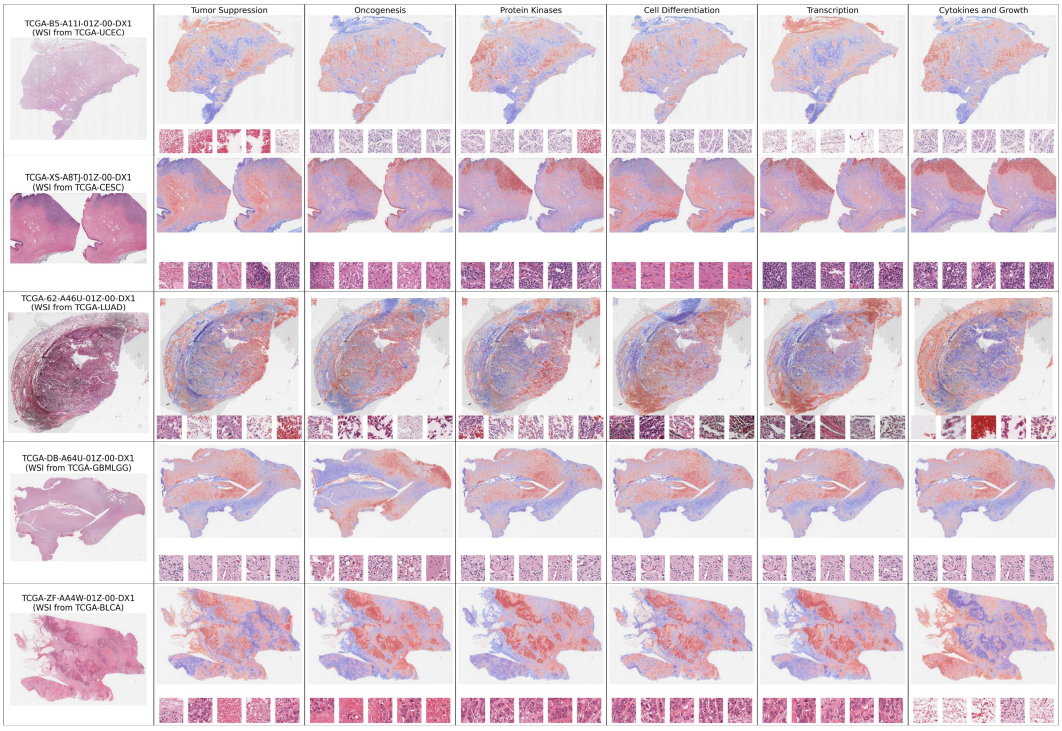

Figure 7: Visualization of cross-attention weights in low-risk cases.

This spatial pattern suggests that tumor suppressor genes (TSGs) are transcriptionally upregulated and broadly active across the tumor tissue, thereby inhibiting malignant proliferation and limiting the invasive and metastatic potential of the tumor. For instance, in a representative CESC case, the patient exhibits molecular characteristics consistent with well-differentiated tumor cells and a tumor-suppressor-driven phenotype. The tumor suppressor pathways show widespread and strong activity, indicating robust TSG signaling that promotes cell-cycle arrest, DNA repair, and apoptosis. In contrast, oncogenic pathways exhibit relatively weak and spatially confined activity, suggesting that the proliferative activity of the tumor cells is tightly regulated.

Overall, these findings highlight a balanced expression of inflammatory and growth-related factors, with no significant pro-inflammatory or pro-angiogenic signals driving tumor progression. The favorable prognosis observed in this patient is likely attributed to the dominant activity of tumor suppressor pathways. This robust tumor-suppressor activity effectively restrains tumor progression, ultimately contributing to the patient's better survival outcome.