LaCONIC: A Label-Aware and Graph-Guided Contrastive Multi-Omics Collaborative Learning Model for Cancer Risk Prediction

Pei Liu, Xiao Liang, Jiawei Luo*.

***Abstract*—Investigating accurate cancer survival prediction models has important clinical value for optimizing therapeutic strategies and improving clinical outcomes. Although an increasing number of models have shifted from relying solely on clinical variables to integrating multi-omics data, there remains insufficient exploitation and integrative utilization of gene regulatory network (GRN) structures and cancer subtype information, which limits the predictive accuracy and biological interpretability of multi-omics models. Therefore, this study proposes label-aware and graph-guided multi-omics collaborative learning framework, termed LaCONIC, to achieve accurate and interpretable cancer survival prediction. To obtain biologically meaningful regulatory features of different molecular types, we first construct a gene regulatory network comprising multiple molecular entities and their relations, and design a heterogeneous graph self-supervised pre-training module to obtain unified graph-based gene representations. To fully leverage multi-omics and multi-modal information, we develop an adaptive cross-omics representation learning module that performs intra-omics modality alignment (i.e., graph and expression modals) and inter-omics representation alignment, thereby achieving synergistic integration of multi-source information. Moreover, to comprehensively capture global cross-omics molecular interactions, we present a graph-guided cross-omics interactive learning module that explicitly encodes gene regulatory network priors within a Transformer-based architecture. Finally, we introduce cancer subtype information to construct a label-aware constraint mechanism that improves predictive performance while enhancing inter-class separability and intra-class consistency of the learned representations. Experiments on multiple cancer datasets show that LaCONIC outperforms existing state-of-the-art methods in various metrics. SHAP-based interpretability analysis on breast cancer further identifies survival-associated regulatory modules and high-risk molecules (e.g., hsa-miR-148a-3p, ARMC1, TMEM242), highlighting its potential for biological mechanism elucidation and prognostic assessment. The source code is available at https://github.com/Liangyushi/LaCONIC.**

***Index Terms*—Cancer survival prediction, Gene regulatory network, Multi-omics integration, Transformer, Heterogeneous graph network.**

I. Introduction

CANCER, as a leading cause of mortality worldwide, is characterized by pronounced biological heterogeneity, and its initiation and progression are driven by perturbations in multi-layered molecular regulatory networks [1]–[3]. In this context, constructing accurate survival prediction models is crucial for risk assessment, prognostic stratification, and individualized treatment.

Early cancer survival prediction models, such as Cox proportional hazards model (CoxPH) [4], LASSO-Cox [5] and Coxnet [6] primarily relied on clinical characteristics (e.g., age, sex, subtype, stage) and classical statistical techniques. These methods offered good interpretability, but depended on manual feature construction and linear assumptions, limiting their ability to capture higher-order nonlinear patterns. With the advent of neural networks, deep survival models such as DeepSurv and DeepHit were proposed. DeepSurv replaced the linear risk function with a multilayer neural network to learn complex risk representations [7], while DeepHit discretized continuous time to directly model the joint distribution of survival time and competing risks [8]. Subsequently, Transformer-based methods (e.g., SurvTRACE [9], SSTrans-STG [10]) further exploited the capacity of Transformers to capture sequences and long-range temporal dependencies [11], thereby improving performance in long-term follow-up and competing-risk settings. However, these approaches relied on manually specified time intervals and remained limited in their ability to capture tumor molecular heterogeneity.

With the rapid development of high-throughput sequencing technologies, the biological landscape of cancer has been systematically profiled across multiple molecular layers, including the genome, transcriptome, epigenome, and proteome [12], [13]. Accumulating evidence has revealed pronounced molecular heterogeneity in cancer and demonstrated that such heterogeneity profoundly influences clinical outcomes [14]–[16]. This indicated that clinical information or single-omics data alone are insufficient to capture the key molecular characteristics required for clinical decision-making. Meanwhile, large public resources such as The Cancer Genome Atlas (TCGA) [17] and the International Cancer Genome Consortium (ICGC) [18] have provided rich multi-omics datasets across diverse cancer types, enabling extensive multi-layer molecular studies. On this basis, numerous multi-omics survival models have been proposed, such as DCAP [19], HFBSurv [20], CAMR [21], and PCLSurv [22]. Most of these methods were built under a continuous-time framework and achieved improved predictive performance and better characterization of cancer molecular heterogeneity compared with clinic feature-based approaches. However, they generally fail to adequately incorporate the topological characteristics of gene regulatory networks and the impact of cancer subtypes on survival heterogeneity, thereby limiting predictive accuracy

Corresponding author: Jiawei Luo. E-mail: luojiawei@hnu.edu.cn.

Pei Liu, Xiao Liang and Jiawei Luo are with the College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China.

Manuscript received xx xx, 2025; revised xx xx, 2025.

and biological interpretability.

In recent years, numerous studies have shown that dysregulation of competing endogenous RNA (ceRNA) networks disrupt miRNA-mediated post-transcriptional regulatory homeostasis, leading to aberrant oncogene activation and tumor suppressor gene silencing, and thereby promoting tumor initiation, progression, and metastasis [23]–[25]. Moreover, different tumor subtypes often exhibit subtype-specific miRNA–ceRNA regulatory modules that are closely associated with patient prognosis, subtype identity, and therapeutic sensitivity [26]–[28]. These observations motivate the joint incorporation of ceRNA network information and cancer subtype annotations into multi-omics survival models, both to more faithfully integrate the regulatory roles of noncoding RNAs and to explicitly preserve subtype-related structural differences during representation learning, thereby enhancing model generalizability and biological consistency.

In this study, we propose LaCONIC, a graph-guided and label-aware multi-omics collaborative learning framework for cancer survival prediction. This framework systematically learn how molecular regulatory networks and cancer subtypes jointly relate to clinical outcomes from a multi-omics perspective, thereby enabling multi-level information integration, survival risk prediction, and biological interpretability. Specifically, a gene regulatory network integrating multiple molecular types and diverse relationships is first constructed to comprehensively characterize complex inter-molecular regulatory patterns. Based on this network, a heterogeneous graph self-supervised pretraining module is developed, which employs node-masking prediction and contrastive learning to generate unified graph-based gene representations that capture both biological and structural information. An adaptive cross-omics representation learning module is then presented, which employs multi-head attention and gating mechanisms to integrate the learned graph-based gene embedding with gene expression data. This module fuses multiple modalities within each omics and performs cross-omics modeling through coordinated components, yielding biologically interpretable latent representations. To explore higher-order dependencies among moleculars, a graph-guided cross-omics interactive learning module is designed to model gene interactions across omics using a Transformer-based structure. Finally, a label-aware constraint mechanism is introduced, integrating subtype classification with multi-level contrastive learning to enhance survival prediction while improving the discriminability and consistency of latent features. Experimental results on multiple cancer datasets demonstrate that LaCONIC outperforms existing methods in various metrics such as Harrell's C-Index. Interpretability analysis combining SHAP and univariate CoxPH on breast cancer identified key survival-associated regulatory modules and biomarkers, providing insights into cancer heterogeneity and targeted therapy.

II. Related work

Cancer survival prediction models aim to estimate the survival probability or recurrence risk of patients based on clinical and multi-omics characteristics, thereby providing a quantitative reference for clinical decision-making. Next, we introduce existing methods from the following three aspects:

A. Based on clinic features

Classical cancer survival prediction models are predominantly built on clinical and pathological features, with the semi-parametric Cox proportional hazards model (CoxPH) as a representative example [4]. On this basis, LASSO-Cox [5] and ElasticNet-Cox (Coxnet) [6] were developed to achieve variable selection and stable modeling via L1/L2 regularization, while Random Survival Forests (RSF) was proposed to capture nonlinear effects and high-order interactions and to handle censored and missing data [29]. With the advancement of deep learning, DeepSurv replaced the linear risk function with a neural network to capture more complex covariate–risk relationships [7], while DeepHit relaxed the proportional hazards assumption by discretizing time and directly modeling the joint distribution of survival time and competing risks [8]. Based on DeepHit, Transformer-based survival models have been proposed to capture global temporal dependencies [30], while SurvTRACE extends this work to model multiple competing events [9], and subsequent semi-supervised frameworks SSTrans-STG improve generalization in small-sample and highly censored settings [10]. Although these advanced transformer-based methods perform well in modeling temporal dynamics, they rely on manually defined time bins whose resolution directly affects performance and interpretability. Moreover, their exclusive reliance on clinical variables makes it difficult to capture the molecular heterogeneity of tumors.

B. Based on multi-omics features

With advances in high-throughput sequencing and integrative multi-omics profiling, survival models have shifted from limited clinical or single-omics features to integrating multi-omics data to better capture tumor molecular heterogeneity. For example, Cox-nnet structurally integrates neural networks and regularization mechanisms, enhancing the ability of multi-omics modeling [31]. DCAP introduces an autoencoder structure to achieve multi-modal feature dimensionality reduction and reconstruction [19]. TransSurv uses a Transformer to fuse histopathology images and genomic features [32]. HFBSurv adopts a bilinear factorization for multi-omics data integration to capture higher-order interactions [20]. FGCNSurv combines factorized bilinear models with graph convolutional networks to enhance prediction accuracy [33]. CAMR introduces a cross-aligned multi-omics representation learning network that generates modal invariant representations to improve robustness [21]. PCLSurv adopts prototype-based contrastive learning to capture high-level semantic relations that effectively distinguishs samples [22]. Although these models can effectively improve the performance, they lack explicit modeling of gene regulatory relationships and their biological interpretability remains limited.

C. GRN-informed

With an increasingly refined understanding of tumor molecular mechanisms, there is a growing demand to explicitly

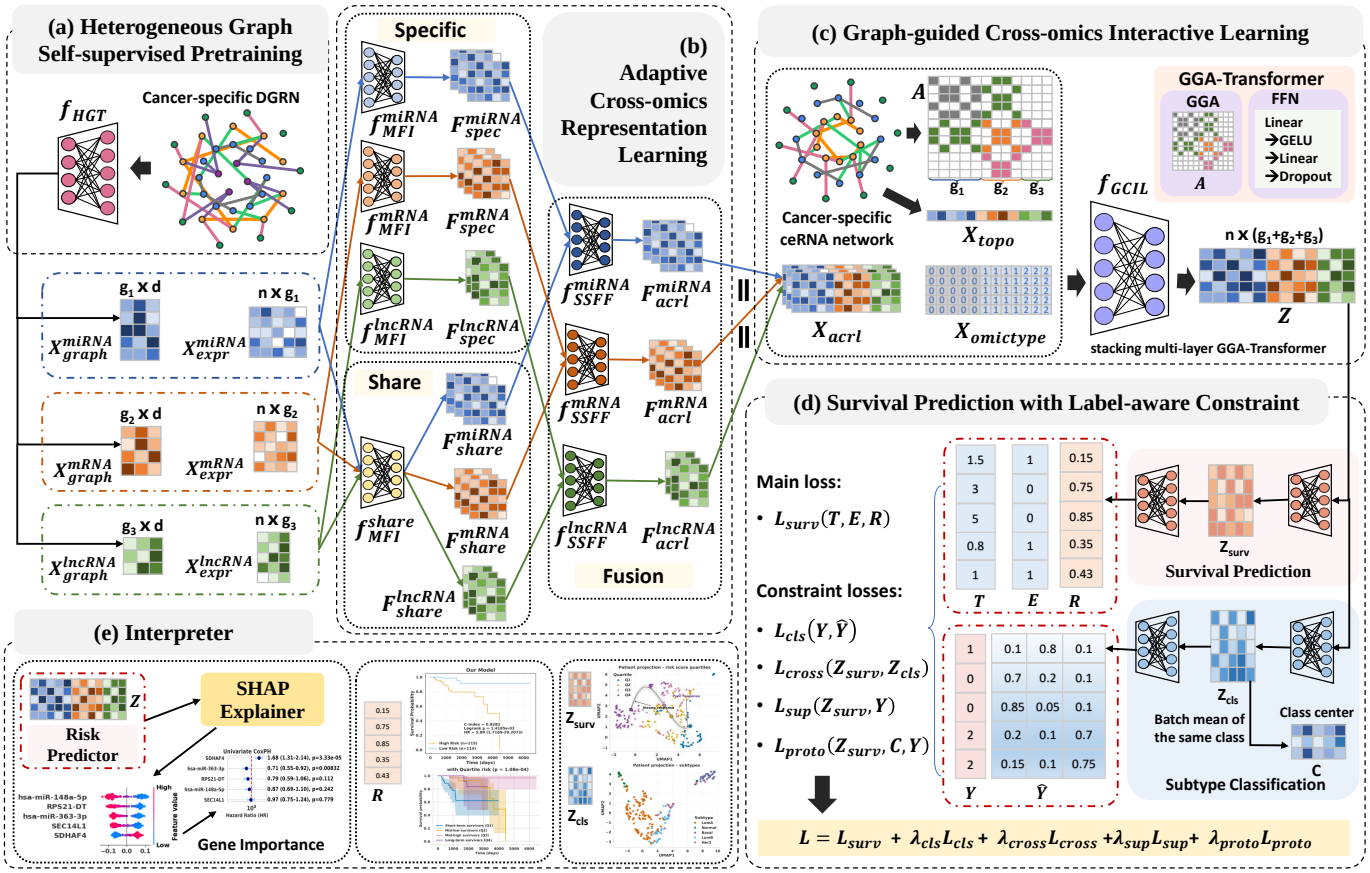

Fig. 1. Overview of the LaCONIC framework. (a) Heterogeneous Graph Self-supervised Pretraining: an HGT-based network learns molecular representations that integrate biological semantics and topological structure. (b) Adaptive Cross-omics Representation Learning: multi-head attention and gating collaboratively model omics-specific and shared features to adaptively align and fuse multi-omics information, yielding robust and biologically interpretable latent representations. (c) Graph-guided Cross-omics Interaction Learning: a graph-guided Transformer models cross-omics gene–gene interactions, capturing cross-layer dependencies and reflecting molecular heterogeneity across patients. (d) Survival Prediction with Label-aware Constraint: survival risk prediction is jointly optimized with subtype classification and multi-level contrastive learning to enhance the discriminability and consistency of the latent feature space. (e) Interpreter: SHAP and univariate Cox regression are used to identify survival-associated key genes and support downstream analyses such as risk stratification.

encode gene regulatory structures as prior knowledge in survival prediction models, thereby enhancing discriminative performance while providing traceable biological interpretations. For example, Sun et al. introduced a graph Laplacian penalty into the Cox framework to smooth the coefficients of neighboring genes [34]. Liu et al. inferred sample-level pathway activities based on pathway topology to strengthen mechanistic relevance [35]. Within deep learning–based survival frameworks, KEGG pathway structures have been encoded as intermediate representations to enable end-to-end interpretable learning [36], while GRN priors have been integrated into multi-omics training to identify key regulatory nodes [37]. Although these approaches help improve robustness and biological interpretability at the gene level, the utilization of post-transcriptional regulatory networks (e.g., ceRNA) for risk stratification remains limited.

Overall, these methods represent a shift in survival prediction from feature-driven to network-driven modeling, moving from single clinical data to multi-omics integration and structured modeling. Nevertheless, existing models remain insufficient in jointly characterizing multi-omics features, gene regulatory network structures, and cancer subtypes. This makes them difficult to fully capture the molecular heterogeneity of cancer and limit their predictive performance and biological interpretability. Thus, a unified framework that jointly integrates these components to model complex molecular interactions is critically needed.

III. Materials and Methods

The LaCONIC collaboratively integrates multi-omics expression data, gene regulatory network structure, and cancer subtype into a unified deep survival prediction framework. The overall framework of LaCONIC is shown in Figure 1.

A. Dataset

We first systematically downloaded and curated multi-source relational data for miRNAs, mRNAs, lncRNAs, and disease entities (see Supplementary Table S1 for details), and constructed a large-scale disease gene regulatory network (DGRN) containing seven molecular associations: miRNA-mRNA, miRNA-lncRNA, mRNA-mRNA, mRNA-disease, miRNA-disease, lncRNA-disease, and miRNA-miRNA. Among these, the miRNA-miRNA associations were

TABLE I
STATISTICAL OVERVIEW OF MULTI-OMICS, CERNA NETWORK AND CLINICAL DATA ACROSS CANCER TYPES

Cancer	Abbr.		Sample		miRNA	mRNA	lncRNA	Event	Censored	Label	ceRNA
Breast invasive carcinoma	BRCA		1143		116	4129	141	103	1040	5	561,289
Cervical squamous cell carcinoma and endocervical adenocarcinoma	CESC		255		78	1849	120	47	208	2	127,028
Head and Neck squamous cell carcinoma	HNSC		461		118	1537	55	125	336	2	95,099
Sarcoma	SARC		224		174	2400	90	69	155	4	286,299
Uterine Corpus Endometrial Carcinoma	UCEC		494		70	2931	165	54	440	4	259,070
Colon adenocarcinoma	COAD	COAD_ESCA _READ_STAD	312	921	69	1554	43	179	742	6	91,982
Esophageal carcinoma	ESCA		167								
Rectum adenocarcinoma	READ		107								
Stomach adenocarcinoma	STAD		335								
Glioblastoma multiforme	GBM	GBM_LGG	142	638	193	6285	376	216	422	3	1,280,772
Brain Lower Grade Glioma	LGG		496								

inferred following the previous research [38], [39]. By comprehensively considering the regulatory effects of common target genes and the influence of protein-protein interaction patterns on miRNA synergy, the high-confidence functional synergistic relationships between miRNAs were constructed.

Then the R package TCGAbiolinks was adopted to obtain survival data and miRNA/mRNA expression profiles for 11 cancer types from TCGA. Based on subtype label overlap, we constructed 5 single-cancer and 2 pan-cancer datasets (Table I). From the large DGRN resources, we further extracted cancer-specific DGRN and ceRNA networks to capture differences in regulatory structure across cancers (network statistics are in Supplementary Table S2). To focus on cancer-relevant biomarkers, we adopted disease-specific survival (DSS) events and times as the final survival labels instead of overall survival (OS). High-confidence subtype labels were derived by intersecting three authoritative sources: (a) subtypes defined in [40]; (b) clinical subtype data from the TCGA Pan-Cancer Atlas (PanCanAtlas) via cBioPortal [41]; and (c) molecular subtype annotations from PanCanAtlas on the UCSC Xena platform [42]. This multi-source integration yielded a unified multi-cancer dataset comprising expression profiles, clinical information, and regulatory networks. More details on data preprocessing are in Supplementary Material Section 1.

B. Heterogeneous Graph Self-supervised Pretraining

To fully capture the complex semantic relationships among miRNAs, mRNAs, lncRNAs, and diseases, we design a heterogeneous graph self-supervised pretraining module based on the Heterogeneous Graph Transformer (HGT) [43]. This module jointly optimizes masked node prediction and contrastive learning to learn unified node embeddings across omics and relation types in an unlabeled setting, providing high-quality network feature initialization. Concretely, given a heterogeneous graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with multiple node and edge types, we initialize type-specific embeddings $E^t\text{HGT}$ for each node type $t \in \{\text{miRNA}, \text{mRNA}, \text{lncRNA}, \text{disease}\}$ using an independent embedding network $\text{Embedding}^t\text{HGT}$, and feed them into a heterogeneous graph encoder Encoder composed of two HGTConv layers to obtain the final node representations $H = f_{\text{HGT}}(\mathcal{V}, \mathcal{E})$.

$$\begin{aligned} E^t_{\text{HGT}} &= \text{Embedding}^t_{\text{HGT}}(\mathcal{V}^t) \in \mathbb{R}^{n_t \times d} \\ E &= \{E^t_{\text{HGT}} \mid t \in \mathcal{T}\} \\ H &= \text{Encoder}(E, \mathcal{E}) \\ &= \text{HGTConv}_2\big(\sigma(\text{HGTConv}_1(E, \mathcal{E})), \mathcal{E}\big) \\ H &= \{H^t_{\text{HGT}} \mid t \in \mathcal{T}\}, \quad H^t_{\text{HGT}} \in \mathbb{R}^{n_t \times d_{\text{HGT}}} \end{aligned} \tag{1}$$

where n_t denotes the number of nodes of type t, d is the initial node feature dimension, and d_{HGT} is the output embedding dimension. The function $\sigma(\cdot)$ denotes the GELU activation. HGTConv is the core operator of HGT, which captures cross-type interactions via meta-relation–specific attention. The output H contains the embeddings of all node types.

To encourage LaCONIC to learn local context-aware semantics, we randomly mask a proportion r of nodes for each node type ($r = 0.15$) and require the model to reconstruct their original indices from the graph context. For each node type t, we define an index decoder $f^t_{\text{dec}}(H^t_{\text{HGT}}) = \text{Linear}(H^t_{\text{HGT}})$. A cross-entropy loss over masked nodes is then used as the masked node prediction objective.

$$\mathcal{L}_{\text{mask}} = \sum_{t \in \mathcal{T}} \text{CE}\big(\hat{I}^t, I^t_{\text{masked}}\big) \tag{2}$$

where $\hat{I}^t = f^t_{\text{dec}}(H^t_{\text{HGT}}) \in \mathbb{R}^{n_t}$ is the reconstructed probability distribution over node indices of type t, and I^t_{masked} denotes the ground-truth index labels for the masked nodes.

To enforce global consistency of node embeddings across different views, we adopt a SimCLR [44] contrastive learning objective. Specifically, we generate two randomly masked views for each node and obtain the corresponding representations $\{z^1_i, z^2_i\}_{i=1}^n$. The contrastive loss is defined as:

$$\mathcal{L}_{\text{con}} = -\frac{1}{2n} \sum_{i=1}^{n} \log \frac{\exp\big(\text{sim}(z^1_i, z^2_i)/\tau\big)}{\sum_{j \neq i} \exp\big(\text{sim}(z^1_i, z^2_j)/\tau\big)} \tag{3}$$

where $\text{sim}(a, b) = a^\top b$ denotes the dot-product similarity and τ is a temperature parameter. In this task, all node types are jointly included in the contrastive set, i.e., n denotes the total number of nodes across all types, which encourages global consistency modeling across different biological entity types.

We then jointly optimize these two self-supervised objectives to obtain the overall pretraining loss:

$$\mathcal{L}_{\text{pre}} = \mathcal{L}_{\text{mask}} + \frac{1}{2}\mathcal{L}_{\text{con}} \tag{4}$$

Through this pretraining strategy, we obtain the high-quality initial molecular graph features $X^o_{\text{graph}} = H^o_{\text{HGT}} \in \mathbb{R}^{g_o \times d_{\text{HGT}}}$, where g_o denotes the number of genes (molecular entities) in omics $o \in \{\text{miRNA}, \text{mRNA}, \text{lncRNA}\}$.

C. Adaptive Cross-omics Representation Learning

To fully exploit the complementarity and consistency of multi-omics, multi-modal data, LaCONIC introduces an Adaptive Cross-omics Representation Learning (ACRL) module that performs intra-omics multi-modal fusion and inter-omics alignment to obtain robust, biologically interpretable latent representations. ACRL comprises three components: an omics-specific encoder, an omics-shared encoder, and a specific–shared fusion layer. Both encoders share a Multi-view Feature Integration (MFI) backbone: the omics-specific encoder uses separate MFI networks for each omics to capture modality-specific patterns, whereas the shared encoder uses a common MFI network across omics to align their latent spaces and learn cross-modal commonalities. The fusion layer then employs a gating mechanism to combine specific and shared features into representations that preserve omics specificity while enforcing cross-omics consistency.

1) Omics-specific encoder: The core component of omics-specific encoder is the Multi-view Feature Integration (MFI) network, whose key idea is to jointly learn intra-omic multi-modal representations from gene expression and graph-based features via a gating mechanism and multi-head attention. On this basis, it instantiates an independent, parameterized MFI network f^o_{MFI} for each omics o, ensuring that LaCONIC can learn omics-specific feature representation for each omic. Specifically, for the o-th omics, the inputs consist of a gene expression matrix $X^o_{\text{expr}} \in \mathbb{R}^{n \times g_o}$ and a gene-level graph feature matrix $X^o_{\text{graph}} \in \mathbb{R}^{g_o \times d_{\text{in}}}$, where n is the number of samples and $d_{\text{in}} = d_{\text{HGT}}$ denotes the initial graph feature dimension. The MFI module $f^o_{\text{MFI}}(X^o_{\text{expr}}, X^o_{\text{graph}})$ is composed of the following four components:

Intra-modal: feature projection. For each omics o, we first replicate the precomputed graph-based gene representations across all samples, yielding $X^o_{\text{graph}} \in \mathbb{R}^{n \times g_o \times d_{\text{in}}}$. Similarly, the expression matrix is broadcast along the gene dimension to $X^o_{\text{expr}} \in \mathbb{R}^{n \times g_o \times 1}$ and then passed through a gene-wise mapping layer $f^{\text{expr}}_{\text{map}}$ to project it into the same feature space as the graph representation $X^o_{\text{expr}} = f^{\text{expr}}_{\text{map}}(X^o_{\text{expr}}) \in \mathbb{R}^{n \times g_o \times d_{\text{in}}}$.

The resulting expression features X^o_{expr} are then fed into an expression-specific projection layer $f^{\text{expr}}_{\text{proj}}$ and a shared projection layer $f^{\text{share}}_{\text{proj}}$ to obtain expression-specific representations E^o_p and shared representations E^o_s, seperately:

$$\begin{aligned} E^o_p &= f^{\text{expr}}_{\text{proj}}(X^o_{\text{expr}}) \in \mathbb{R}^{n \times g_o \times d_{\text{out}}} \\ E^o_s &= f^{\text{share}}_{\text{proj}}(X^o_{\text{expr}}) \in \mathbb{R}^{n \times g_o \times d_{\text{out}}} \end{aligned} \tag{5}$$

Analogously, the replicated graph features X^o_{graph} are processed by a graph-specific projection layer $f^{\text{graph}}_{\text{proj}}$ and the same shared projection layer $f^{\text{share}}_{\text{proj}}$, yielding graph-specific and shared graph representations as:

$$\begin{aligned} G^o_p &= f^{\text{graph}}_{\text{proj}}(X^o_{\text{graph}}) \in \mathbb{R}^{n \times g_o \times d_{\text{out}}} \\ G^o_s &= f^{\text{share}}_{\text{proj}}(X^o_{\text{graph}}) \in \mathbb{R}^{n \times g_o \times d_{\text{out}}} \end{aligned} \tag{6}$$

where $f^{\text{expr}}_{\text{map}}$, $f^{\text{expr}}_{\text{proj}}$, $f^{\text{graph}}_{\text{proj}}$, and $f^{\text{share}}_{\text{proj}}$ are all implemented as single-layer blocks consisting of a linear transformation, Layer Normalization (LN), and a GELU activation.

Inter-modal: shared features fusion. To balance the contributions of expression and graph features to the shared representation, f^o_{MFI} first concatenates the two shared components, and feeds $S^o = [E^o_s \,\|\, G^o_s] \in \mathbb{R}^{n \times g_o \times 2d_{\text{out}}}$ into a channel-wise gating layer f^o_{gated} to achieve adaptive fusion: To balance the contributions of expression and graph features to the shared representation, f^o_{MFI} first concatenates the two shared components to form $S^o = [E^o_s \,\|\, G^o_s] \in \mathbb{R}^{n \times g_o \times 2d_{\text{out}}}$, and then feeds S^o into a channel-wise gating layer f^o_{gated} to achieve adaptive fusion as:

$$\begin{aligned} G^o_{\text{gate}} &= f^o_{\text{gated}}(S^o) = \sigma\big(W^o S^o\big) \in \mathbb{R}^{n \times g_o \times d_{\text{out}}} \\ Z^o_{\text{share}} &= G^o_{\text{gate}} \odot E^o_s + (1 - G^o_{\text{gate}}) \odot G^o_s \in \mathbb{R}^{n \times g_o \times d_{\text{out}}} \end{aligned} \tag{7}$$

where $\sigma(\cdot)$ denotes the sigmoid function, G^o_{gate} is the learned gating weight, and $\odot$ denotes element-wise multiplication.

Intra-inter modal: multi-modal feature fusion. We first combine the shared representation with the modality-specific components and normalize the results to obtain enhanced expression and graph features. These are then concatenated to form a unified multi-view representation H^o as:

$$\begin{aligned} E^o &= \text{LN}^o_{\text{expr}}\big(E^o_p + Z^o_{\text{share}}\big) \in \mathbb{R}^{n \times g_o \times d_{\text{out}}} \\ G^o &= \text{LN}^o_{\text{graph}}\big(G^o_p + Z^o_{\text{share}}\big) \in \mathbb{R}^{n \times g_o \times d_{\text{out}}} \\ H^o &= [\, G^o \,\|\, E^o \,] \in \mathbb{R}^{n \times g_o \times 2d_{\text{out}}} \end{aligned} \tag{8}$$

Intra-omics: gene interaction learning. To capture higher-order gene dependencies within each omics, f^o_{MFI} applies multi-head self-attention MHA to H^o, followed by residual connection and normalization, yielding the final omics-specific integrated representation F^o_{spec} for omic o:

$$\begin{aligned} H^o_{\text{attn}} &= \text{MHA}^o(H^o, H^o, H^o) \in \mathbb{R}^{n \times g_o \times 2d_{\text{out}}} \\ F^o_{\text{spec}} &= \text{LN}^o_{\text{res}}(H^o_{\text{attn}} + H^o) \in \mathbb{R}^{n \times g_o \times 2d_{\text{out}}} \end{aligned} \tag{9}$$

2) Omics-shared encoder: This encoder is designed to align latent representation distributions across different omics, thereby capturing cross-modal consistency. Similar to the omics-specific layer, it is also built upon the MFI architecture, but adopts a parameter-sharing strategy. In other words, all omics share the same MFI network $f^{\text{share}}_{\text{MFI}}$, which enforces distributional alignment and semantic consistency in a common feature space. Formally, the shared representations for o omic can be written as:

$$F^o_{\text{share}} = f^{\text{share}}_{\text{MFI}}\big(X^o_{\text{expr}}, X^o_{\text{graph}}; \theta_{share}\big) \in \mathbb{R}^{n \times g_o \times 2d_{\text{out}}} \tag{10}$$

where θ_{share} is a shared parameter for all omics.

3) Specific-shared fusion layer: This layer is designed to integrate representations from the omics-specific and omics-shared encoders, yielding final features that jointly preserve modality-specific information and cross-omics consistency. Its core component is the Specific-Shared Feature Fusion (SSFF) module. SSFF first performs channel-wise gated fusion between specific and shared representations, and then applies Dropout and Layer Normalization (LN) to obtain a stable and fused omics-level representation. Formally, given the omics-specific and shared features for the o-th omics, the fusion process implemented by f^o_{SSFF} is defined as:

$$\begin{aligned} G^o &= \sigma^o\big(W^o[F^o_{\text{spec}} \,\|\, F^o_{\text{share}}]\big) \in \mathbb{R}^{n\times g_o\times 2d_{\text{out}}} \\ F^o_{\text{ACRL}} &= \text{LN}^o\Big(\text{Dropout}^o\big(G^o \odot F^o_{\text{spec}} \\ &\quad + (1-G^o)\odot F^o_{\text{share}}\big)\Big) \in \mathbb{R}^{n\times g_o\times 2d_{\text{out}}} \end{aligned} \tag{11}$$

This mechanism adaptively modulates the relative contributions of specific and shared information for each feature channel, enabling fine-grained information balancing.

D. Graph-guided Cross-omics Interactive Learning

The Graph-guided Cross-omics Interactive Learning (GCIL) module aims to learn context-dependent and cross-modal gene interactions under the guidance of the ceRNA network, enabling the model to capture key regulatory genes and their global impact on cancer prognosis. Before feeding data into GCIL, we first concatenate miRNA, mRNA, and lncRNA features along the gene dimension to obtain $X_{\text{ACRL}} = \big[F^{\text{miRNA}}_{\text{ACRL}} \,\|\, F^{\text{mRNA}}_{\text{ACRL}} \,\|\, F^{\text{lncRNA}}_{\text{ACRL}}\big] \in \mathbb{R}^{n\times g\times 2d_{\text{out}}}$, where $g = g_{\text{miRNA}} + g_{\text{mRNA}} + g_{\text{lncRNA}}$ denotes the total number of genes. Each gene is additionally assigned an omics-type label $T = [t_{\text{miRNA}}, t_{\text{mRNA}}, t_{\text{lncRNA}}]$, $t_o \in \{0,1,2\}$, which is used to construct the omics-type feature embedding $X_{\text{omictype}} \in \mathbb{R}^{n\times g}$. Based on the ceRNA network adjacency matrix $A \in \mathbb{R}^{g\times g}$, we further compute the betweenness centrality of each gene to obtain a topological feature vector $X_{\text{topo}} \in \mathbb{R}^{g\times 1}$ (Specific calculation process can be found in Supplementary Material Section 2). Thus, given X_{ACRL}, X_{omictype}, X_{topo}, and A, the GCIL module produces sample-level representations as:

$$Z = f_{\text{GCIL}}\big(X_{\text{ACRL}}, X_{\text{omictype}}, X_{\text{topo}}, A\big) \in \mathbb{R}^{n\times 2d_{\text{out}}} \tag{12}$$

In the following, we detail the architecture of f_{GCIL}.

1) Graph-guided attention mechanism: To explicitly model regulatory dependencies among genes, GCIL introduces a Graph-Guided Attention (GGA) mechanism, which augments standard multi-head self-attention with structural priors from the ceRNA adjacency matrix A. Formally, given input features $X \in \mathbb{R}^{n\times g\times 2d_{\text{out}}}$ and adjacency matrix $A \in \mathbb{R}^{g\times g}$, the attention weights are defined as:

$$S = \text{Dropout}\left(\text{softmax}\left(\frac{QK^\top}{\sqrt{d_{\text{head}}}} + \alpha A\right)\right), \tag{13}$$

where $Q = \text{reshape}(W_qX), K = \text{reshape}(W_kX), V = \text{reshape}(W_vX) \in \mathbb{R}^{n\times h\times g\times (2d_{\text{out}}/h)}$ are the query, key, and value tensors for h attention heads, $W_q, W_k, W_v \in \mathbb{R}^{2d_{\text{out}}\times 2d_{\text{out}}}$ are learnable projection matrices, $d_{\text{head}} = 2d_{\text{out}}/h$ is the per-head dimensionality, and α is a learnable bias coefficient that controls the strength of the graph structural prior. GGA can guide models to focus more on known biological regulatory connections while retaining the ability to explore new potential connections. Then, the graph-guided contextual features are obtained by aggregating the value tensor with the attention weights, followed by a linear projection as:

$$X_{\text{att}} = W_{\text{att}}(SV) \in \mathbb{R}^{n\times g\times 2d_{\text{out}}}, \tag{14}$$

where $W_{\text{att}} \in \mathbb{R}^{2d_{\text{out}}\times 2d_{\text{out}}}$ is a learnable parameter matrix.

2) GGA-Transformer backbone encoder: By integrating GGA with Transformer, GCIL constructs a GGA-Transformer backbone consisting of a GGA mechanism and a two-layer feed-forward network (FFN) for nonlinear feature transformation. GCIL models cross-omics gene interactions by stacking multiple GGA-Transformer encoder layers, with residual connections and layer normalization applied in each layer to stabilize training and enhance representational capacity.

$$\begin{aligned} X &= \text{LN}_1\big(X + \text{GGA}(X)\big) \\ X &= \text{LN}_2\big(X + \text{FFN}(X)\big) \in \mathbb{R}^{n\times g\times 2d_{\text{out}}} \end{aligned} \tag{15}$$

To capture semantic differences arising from omics type and topological context, we further introduce two types of embeddings:

$$X = X_{\text{ACRL}} + E_{\text{omictype}} + E_{\text{topo}} \in \mathbb{R}^{n\times g\times 2d_{\text{out}}}, \tag{16}$$

where $E_{\text{omictype}} = \text{Embedding}_{\text{omictype}}(X_{\text{omictype}}) \in \mathbb{R}^{g\times 2d_{\text{out}}}$ is the omics-type embedding, and $E_{\text{topo}} = \text{LN}_{\text{topo}}(W_{\text{topo}}X_{\text{topo}}) \in \mathbb{R}^{g\times 2d_{\text{out}}}$ is the linearly projected and normalized topological embedding, with W_{topo} being a learnable weight matrix.

Finally, GCIL applies a gated pooling scheme that combines average and max pooling across the gene dimension to obtain sample-level representations as:

$$Z = \sigma(\alpha)\cdot \text{avg}(X) + \big(1-\sigma(\alpha)\big)\cdot \max(X) \in \mathbb{R}^{n\times 2d_{\text{out}}} \tag{17}$$

where α is a learnable scalar, $\sigma(\cdot)$ is the sigmoid function, and $\text{avg}(\cdot)$ and $\max(\cdot)$ denote average and max pooling over the gene dimension, respectively.

E. Survival Prediction with Label-aware Constraint

LaCONIC jointly optimizes survival prediction and label-aware constraint to maintain high survival prediction accuracy while learning structurally coherent and biologically interpretable multi-omics representations.

1) Survival prediction: This constitutes the core optimization objective of LaCONIC and is used to learn continuous representations that reflect patient-specific prognostic risk. Given the fused multi-omics representation Z, a survival head f_{surv} based on multilayer perceptron (MLP) is employed to map Z nonlinearly to scalar risk scores $R = f_{\text{surv}}(Z) = \{r_1, \ldots, r_i, \ldots, r_n\}$, where r_i is the risk score of the i-th sample. To balance the contributions of events and censored observations, we design a weighted Cox loss as:

$$\mathcal{L}_{\text{surv}} = -\frac{1}{\sum_{i=1}^n \omega_i\delta_i}\sum_{i=1}^n \omega_i\delta_i\left(r_i - \log\sum_{j:\,t_j\geq t_i} e^{r_j}\right) \tag{18}$$

where $\delta_i \in \{0,1\}$ is the event indicator ($\delta_i = 1$ if the event occurs), t_i is the observed survival time, and ω_i is the sample weight. In practice, we set $\omega_i = \frac{n}{n_{\delta=1}}$ for event samples ($\delta_i = 1$) and $\omega_i = 1$ otherwise. This weighting scheme upweights rare event cases, increasing the model's sensitivity to failure events and stabilizing gradient estimation during training.

2) Label-aware constraint: This constraint includes a subtype classification loss and three contrastive objectives, which encourages both consistency and discriminability of the learned representations, thereby imposing explicit regularization on the feature space and improving survival prediction.

Subtype classification constraint. The cancer subtype classification task acts as a supervised auxiliary signal that encourages biologically coherent and well-separated subtype representations in the latent space, providing complementary supervision for survival risk modeling. An MLP-based classifier f_{cls} maps the fused representation Z to subtype logits $\hat{Y} = f_{\text{cls}}(Z) \in \mathbb{R}^{n\times K}$, where K is the number of subtypes. To emphasize hard examples and alleviate class imbalance, we adopt the Focal Loss [45]:

$$\mathcal{L}_{\text{cls}} = -\frac{1}{n}\sum_{i=1}^{n} \alpha_{y_i}(1-p_{y_i})^{\gamma}\log p_{y_i} \tag{19}$$

where y_i is the ground-truth of sample i, $p_{y_i} = \mathrm{Softmax}(\hat{y}_i)_{y_i}$ is the predicted probability of the true class, and $\alpha = 1$, $\gamma = 2.0$ are the class-balancing and focusing parameters.

Multi-level contrastive constraint. To further enforce the consistency between survival representations and subtype features, this constraint introduces three following contrastive losses to jointly enhance the global consistency and local discriminability of the representation space.

· Cross-task feature alignment loss. This term aligns the survival and classification representations. Given two task-specific embeddings $Z_{\text{surv}} = \{z_{\text{surv},i}\}$ and $Z_{\text{cls}} = \{z_{\text{cls},i}\}$ for each sample, we maximize their mutual information via a bidirectional InfoNCE loss [46] as:

$$\mathcal{L}_{\text{cross}} = -\frac{1}{2n}\sum_{i=1}^{n}\Big[\log\frac{\exp(z_{\text{surv},i}\cdot z_{\text{cls},i}/\tau_{\text{cross}})}{\sum_{j=1}^{n}\exp(z_{\text{surv},i}\cdot z_{\text{cls},j}/\tau_{\text{cross}})} + \log\frac{\exp(z_{\text{cls},i}\cdot z_{\text{surv},i}/\tau_{\text{cross}})}{\sum_{j=1}^{n}\exp(z_{\text{cls},i}\cdot z_{\text{surv},j}/\tau_{\text{cross}})}\Big] \tag{20}$$

where τ_{cross} denotes the temperature, and Z_{surv} and Z_{cls} are the ℓ_2-normalized penultimate-layer embeddings of the survival head f_{surv} and classification head f_{cls}, respectively.

· Label-supervised contrastive loss. This loss imposes a label-informed discriminative structure on the survival embedding space by pulling same-subtype samples together and pushing different subtypes apart. For each anchor i, given normalized survival embeddings $z_{\text{surv},i}$, ground-truth labels y_i, and the positive set $P(i) = \{p \neq i \mid y_p = y_i\}$, the supervised contrastive loss [47] is defined as:

$$\mathcal{L}_{\text{sup}} = -\frac{1}{n}\sum_{i=1}^{n}\frac{1}{|P(i)|}\sum_{p\in P(i)}\log\frac{\exp(z_{\text{surv},i}\cdot z_{\text{surv},p}/\tau_{\text{sup}})}{\sum_{a\neq i}\exp(z_{\text{surv},i}\cdot z_{\text{surv},a}/\tau_{\text{sup}})} \tag{21}$$

· Prototype-based contrastive loss. This loss is adopted to enhance the stability of decision boundaries and mitigate cross-batch drift. For each subtype k, LaCONIC maintains a prototype vector c_k in the classification embedding space, initialized from the output of f_{cls} and updated via Exponential Moving Average (EMA) [48]. Let $B_k = i \mid y_i = k$ be the set of samples of subtype k in the current mini-batch, their mean embedding is defined as:

$$\tilde{c}_k = \frac{1}{|B_k|}\sum_{i\in B_k} z_{\text{cls},i} \tag{22}$$

Then the global prototype is updated by $c_k \leftarrow \beta\, c_k + (1-\beta)\,\tilde{c}_k$ with momentum $\beta \in [0,1)$.

Given a survival embedding $z$surv, i and its class prototype c_{y_i} as the positive (all other prototypes as negatives), the prototype-based contrastive loss [49] is defined as:

$$\mathcal{L}_{\text{proto}} = -\frac{1}{n}\sum_{i=1}^{n}\log\frac{\exp(z_{\text{surv},i}\cdot c_{y_i}/\tau_{\text{proto}})}{\sum_{k=1}^{K}\exp(z_{\text{surv},i}\cdot c_k/\tau_{\text{proto}})} \tag{23}$$

where τ_{proto} is a temperature parameter and K is the number of subtypes. By combining the above loss terms, the overall objective of LaCONIC is defined as:

$$\mathcal{L} = \mathcal{L}_{\text{surv}} + \lambda_{\text{cls}}\mathcal{L}_{\text{cls}} + \lambda_{\text{cross}}\mathcal{L}_{\text{cross}} + \lambda_{\text{sup}}\mathcal{L}_{\text{sup}} + \lambda_{\text{proto}}\mathcal{L}_{\text{proto}} \tag{24}$$

where λ_{cls}, λ_{cross}, λ_{sup}, and λ_{proto} are the weighting coefficients for the corresponding auxiliary objectives.

F. SHAP-based Gene Importance

To enhance model interpretability, we employ SHAP (SHapley Additive exPlanations) [50] to quantify the contribution of each gene feature to survival prediction. SHAP is a post-hoc, game-theoretic explanation method that computes Shapley values by estimating the average marginal contribution of each feature over all possible feature coalitions, thereby revealing how individual features influence the model output. Concretely, we initialize a `shap.GradientExplainer` using the trained survival head f_{surv} together with the training representations $Z_{\text{train}} \in \mathbb{R}^{n_{\text{train}}\times g}$. We then feed the test representations $Z_{\text{test}} \in \mathbb{R}^{n_{\text{test}}\times g}$ into the explainer to obtain a SHAP value matrix $S \in \mathbb{R}^{n_{\text{test}}\times g}$, where S_{ij} denotes the marginal contribution of gene j to the predicted risk of sample i. Based on S, we compute the mean absolute SHAP value for each gene across all test samples, use this quantity as the gene importance to rank the genes, and then select the top 20 genes for downstream biological validation to investigate their potential regulatory roles in cancer prognosis.

IV. Experiments

A. Experimental setup

We systematically evaluated LaCONIC using multiple complementary survival analysis metrics, including Harrell's concordance index (Harrell's C-Index) to assess ranking concordance of individual survival times, the time-dependent area under the ROC curve integrated over time (iAUC) to quantify discriminative ability at the 25th, 50th, and 75th percentiles of the survival time distribution, the inverse probability of censoring weighting C-Index (IPCW C-Index) to reduce bias

TABLE II
COMPARISON OF DIFFERENT CANCER SURVIVAL PREDICTION MODELS BASED ON HARRELL'S C-INDEX ACROSS SEVEN CANCER DATASETS. BOLDFACE AND UNDERLINING INDICATE THE BEST AND SECOND-BEST PERFORMANCES, RESPECTIVELY.

	Model	BRCA	CESC	HNSC	SARC	UCEC	COAD_ESCA_READ_STAD	GBM_LGG	Mean±STD
Traditional	Coxnet	0.8120	0.7698	0.6117	0.8116	0.6914	0.5242	0.6455	0.6952±0.11
	CoxBoost	0.7549	0.7415	0.6146	0.7387	0.7360	0.5991	0.8385	0.7176±0.08
	RSF	0.7584	0.7245	0.5836	0.7814	0.7343	0.6784	0.8280	0.7269±0.08
Based on discrete time	DeepHit	0.6860	0.7132	0.6830	0.7487	0.7145	0.6921	0.6876	0.7036±0.02
	Transformer Survival	0.7542	0.8566	0.6787	0.8065	0.8399	0.6905	0.7443	0.7673±0.07
	Trans-STG	0.8002	0.8377	0.7061	0.7940	**0.8449**	0.7176	0.8799	0.7972±0.07
	SurvTRACE	0.8071	0.8679	0.6499	0.8367	**0.8449**	0.7158	0.8606	0.7975±0.08
Based on continuous time	DeepSurv	0.6762	0.7887	0.6477	0.4899	0.7079	0.5718	0.6417	0.6463±0.10
	Cox-nnet	0.7402	0.6868	0.4820	0.7889	0.6766	0.6604	0.7587	0.6848±0.10
	DCAP	**0.8282**	0.8038	0.6398	0.8593	0.7162	0.7222	0.8203	0.7700±0.08
	HFBSurv	0.7554	0.8075	0.6477	0.8116	0.6865	0.6692	0.8326	0.7443±0.08
	CAMR	0.8041	0.7642	0.6001	0.8178	0.7186	0.7011	0.8420	0.7497±0.08
	FGCNSurv	0.8091	0.8264	0.6037	0.8166	0.6898	0.6977	0.8655	0.7584±0.10
	PCLSurv	0.8219	0.7434	0.6441	0.8116	0.7393	0.7263	0.8683	0.7650±0.07
	LaCONIC (our)	**0.8282**	**0.8943**	**0.7075**	**0.8844**	0.8218	**0.7603**	**0.8893**	**0.8266±0.07**

TABLE III
LOG-RANK TEST P-VALUES FROM KAPLAN–MEIER (KM) RISK STRATIFICATION FOR ALL MODELS ON THE SEVEN CANCER DATASETS, WHERE UNDERLINING DENOTES NON-SIGNIFICANT RESULTS AND BOLDFACE DENOTES THE MOST SIGNIFICANT SEPARATION.

	Model	BRCA	CESC	HNSC	SARC	UCEC	COAD_ESCA _READ_STAD	GBM_LGG	Mean±STD
Traditional	Coxnet	2.98E-04	1.00E-02	1.25E-01	1.69E-02	1.64E-01	7.85E-01	2.54E-02	0.1610±0.28
	CoxBoost	3.58E-01	2.78E-02	2.83E-01	8.02E-03	4.30E-02	3.65E-02	6.18E-10	0.1080±0.15
	RSF	4.94E-03	5.01E-02	3.53E-01	6.33E-02	1.43E-01	2.78E-05	1.34E-10	0.0877±0.13
Based on discrete time	DeepHit	8.38E-03	9.97E-02	1.83E-01	2.55E-01	1.03E-01	9.47E-04	7.59E-04	0.0930±0.10
	Transformer Survival	3.92E-02	2.55E-02	1.22E-01	9.20E-02	9.24E-03	1.78E-02	7.45E-06	0.0437±0.05
	Trans-STG	6.44E-04	2.99E-03	1.19E-01	7.99E-03	6.32E-03	1.67E-05	1.26E-11	0.0196±0.04
	SurvTRACE	**1.43E-04**	1.49E-02	1.04E-01	2.07E-03	**4.30E-04**	2.77E-04	**2.50E-13**	0.0174±0.04
Based on continuous time	DeepSurv	3.06E-01	2.07E-02	4.92E-02	7.22E-01	1.28E-02	7.18E-02	3.21E-02	0.1736±0.26
	Cox-nnet	6.94E-02	3.82E-02	5.13E-01	2.55E-03	4.66E-01	2.30E-03	2.00E-06	0.1560±0.23
	DCAP	8.48E-04	2.28E-02	1.83E-01	8.00E-04	2.31E-02	**1.37E-06**	7.22E-09	0.0329±0.07
	HFBSurv	2.43E-02	7.00E-03	4.65E-02	2.78E-03	6.35E-02	2.47E-03	9.64E-06	0.0209±0.03
	CAMR	6.83E-03	3.13E-02	5.21E-01	1.42E-02	2.52E-02	3.10E-04	2.63E-09	0.0855±0.19
	FGCNSurv	4.70E-03	4.93E-03	5.55E-01	5.15E-03	5.40E-02	2.57E-03	3.45E-12	0.0894±0.21
	PCLSurv	4.08E-03	1.85E-02	8.74E-02	1.89E-03	1.50E-01	3.69E-04	1.01E-12	0.0375±0.06
	LaCONIC (ours)	1.42E-03	**1.27E-03**	**3.98E-02**	**1.62E-04**	1.05E-03	1.69E-04	7.61E-11	**0.0063±0.01**

from censoring, and Kaplan–Meier (KM) survival curve analysis with median risk–based high/low risk stratification, where group differences were assessed by the log-rank test. To ensure representative and stable multimodal data, we adopted stratified splitting based on joint strata of event status and subtype label, first dividing each dataset into 80% training and 20% test, and then holding out 20% of the training set for validation (64%/16%/20% train/validation/test). More detailed experimental settings are provided in the Supplementary Material Section 3.

B. Evaluation of survival prediction performance

To evaluate the performance of LaCONIC, we compared it against 14 existing methods, including three classical survival models (Coxnet, CoxBoost [51], RSF), four deep discrete-time models (DeepHit, Transformer Survival, Trans-STG, SurvTRACE), and seven deep continuous-time models (DeepSurv, Cox-nnet, DCAP, HFBSurv, CAMR, FGCNSurv, PCLSurv). To ensure a fair comparison, all methods except FGCNSurv and PCLSurv (which were inherently designed to use only miRNA and mRNA) were trained on the same multi-omics inputs. Moreover, except for HFBSurv, CAMR, FGCNSurv, and PCLSurv (which employed a specific encoder for each omics), the remaining baselines took the concatenated three-omics gene expression matrix as input.

1) Survival ranking concordance: We evaluated all methods on seven cancer datasets (five single-cancer and two pan-cancer) and reported their Harrell's C-Index values in Table II, which depicted that LaCONIC achieves the highest C-Index on all datasets except UCEC, with particularly pronounced gains on CESC (0.8943), SARC (0.8844), and the two pan-cancer datasets COAD_ESCA_READ_STAD (0.7603) and GBM_LGG (0.8893), clearly surpassing all baselines. Traditional methods performed markedly worse, especially on complex multi-cancer data. For example, CoxNet attained only a C-Index of 0.5242 on COAD_ESCA_READ_STAD, far below LaCONIC. Discrete-time models improved over traditional approaches but still fall short of LaCONIC. For instance, Trans-STG obtained C-Index values of 0.7940 on SARC and 0.8377 on CESC, compared with 0.8844 and

0.8943 for LaCONIC. Continuous-time models were competitive on some datasets but do not outperform LaCONIC overall. For example, PCLSurv reached 0.8219 on BRCA, close to LaCONIC, yet remained inferior on CESC and GBM_LGG (0.7434 and 0.8683 vs. 0.8943 and 0.8893).These results demonstrated that LaCONIC provides consistently superior survival ranking concordance across both single-cancer and multi-cancer settings.

Additionally, Supplementary Figure S1 summarized iAUC and IPCW C-Index distributions for all models across seven datasets, ordered by decreasing mean performance. LaCONIC achieved the highest mean *iAUC* (0.8563±0.07), exceeding the second-best method SurvTRACE (0.8220±0.09) by 3.43%. In terms of *IPCW C-Index*, LaCONIC (0.7766±0.08) again exhibited the leading performance, improving on Trans-STG (0.7188±0.11) by 5.78% and substantially outperforming conventional methods. These findings indicated that LaCONIC attained high discriminative accuracy and robustness across diverse cancer survival prediction tasks.

2) Kaplan-Meier risk stratification: To further assess risk stratification ability, we performed KM survival analysis for all models on seven cancer datasets. Table III reported the corresponding log-rank test p-values, where smaller values indicated more pronounced separation between risk groups. LaCONIC achieved statistically significant risk stratification ($p<0.05$) on all datasets, with the smallest p-values observed on CESC (p=1.27E-03), HNSC (p=3.98E-02) and SARC (p=1.62E-04), and its mean p-value (0.0063±0.01) was markedly lower than that of any baselines. In contrast, traditional models and other deep approaches often yielded relatively large p-values on several datasets (e.g., HNSC and UCEC), indicating limited risk stratification capability. Supplementary Figure S2 further illustrated the KM survival curves of all models on the SARC dataset. LaCONIC produced the clearest separation between high and low risk groups, with the high-risk group's survival dropping rapidly and an HR of 17.81, indicating nearly 18-fold higher mortality risk and demonstrating clinically meaningful risk stratification.

C. Evaluation of the correlation between clinical covariates and predicted risk

To examine the correlation between LaCONIC-predicted risk and clinical covariates and assess its independent prognostic value, we performed CoxPH regression on each cancer dataset. For each dataset, we assembled survival time, event status, clinical variables (age, sex, stage, subtype, etc.), and the LaCONIC risk score. Continuous variables (e.g., age, risk score) were z-score standardized and categorical variables (e.g., sex, stage) were one-hot encoded. Samples with missing or implausible values were excluded. Univariable and multivariable CoxPH were fitted using the `lifelines` package [52]. For each covariate, we estimated the hazard ratio (HR) and its 95% confidence interval (CI) in the univariable model. Multivariable CoxPH were used to assess whether the LaCONIC risk score preserved prognostic significance after adjustment for clinical factors.

Figures 2 and Supplementary Figure S3 summarized the univariable and multivariable Cox regression results across all cancer datasets. The LaCONIC risk score consistently emerged as a significant and independent prognostic factor. In BRCA, the risk score was strongly associated with survival in both univariable (HR = 3.60, p = 1.59×10^{-4}) and multivariable (HR = 4.49, p = 1.25×10^{-4}) models, whereas clinical variables such as subtype and stage were not significant, indicating that the predicted risk alone captured major survival heterogeneity. Similar patterns were observed in other cancers: in univariable analyses, the HRs for the predicted risk in CESC, SARC, and GBM_LGG were 3.70, 2.98, and 5.81, respectively (all $p<0.05$), exceeding those of conventional clinical factors. After adjustment for age, sex, subtype, and other covariates, the LaCONIC risk score remained statistically significant ($p<0.05$), with HRs of 3.91, 7.81, and 6.86 in CESC, SARC, and GBM_LGG, respectively, thereby underscoring its role as a key prognostic determinant in survival analysis.

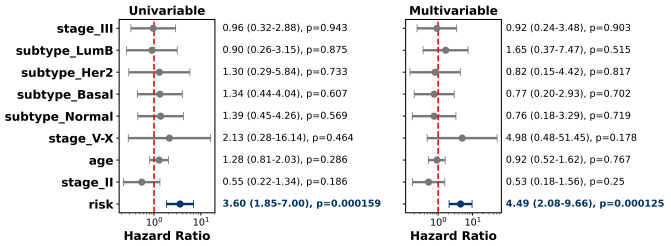

Fig. 2. Univariable and multivariable Cox proportional hazards analysis of LaCONIC-predicted risk and clinical covariates on the BRCA dataset. The forest plot displays the hazard ratio (HR), 95% confidence interval (CI), and p-value for each covariate.

D. Interpretability analysis

To investigate the interpretability of LaCONIC in survival risk prediction, we took the BRCA dataset as an example and conducted a systematic analysis. First, we computed SHAP-based importance to quantify the contribution of each gene to the predicted risk and selected the top 20 genes with the largest impact on the model output (Figure 3a). We then performed univariable CoxPH regression for these genes and identified 14 genes significantly associated with survival ($p<0.05$, Figure 3b). By comparing gene expression levels, SHAP values and the direction of risk in the CoxPH, we found that the correlation between expression and SHAP values was fully consistent with the direction of HR (Figure 3c). Specifically, genes with HR>1 (e.g., hsa-miR-365a-3p, TMEM242, ARMC1) showed a positive correlation between expression and SHAP values, indicating that higher expression increased survival risk, whereas genes with HR<1 (e.g., hsa-miR-148a-3p, hsa-miR-363-3p, GSTM4) showed a negative correlation, suggesting a protective effect of high expression.

1) Analysis of survival risk and gene expression patterns: We stratified patients into four groups according to the quartiles of the LaCONIC-predicted risk scores: short-term survivors (Q1, >75%), mid-low survivors (Q2, 50–75%), mid-high survivors (Q3, 25–50%), and long-term survivors (Q4, <25%). We then analyzed survival outcomes and gene expression patterns across these groups. The KM curves in Figure 3d

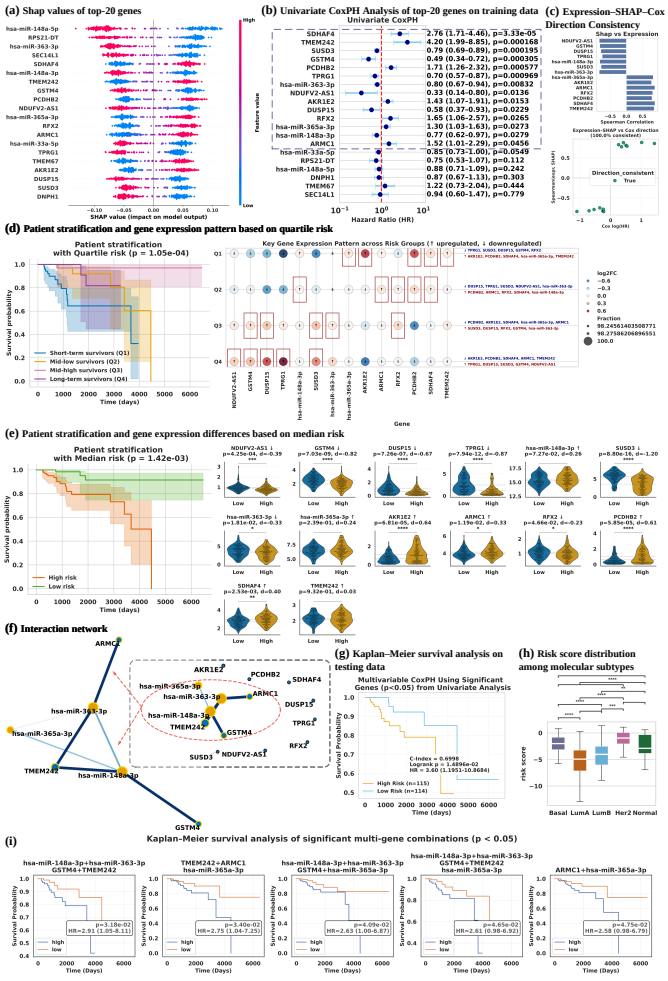

Fig. 3. SHAP-based interpretability analysis of the LaCONIC model.

showed significant survival differences (p=1.05×10^{-4}), indicating that our model effectively distinguished high-risk and low-risk patients: genes such as hsa-miR-148a-3p, AKR1E2, and PCDHB2 were upregulated in the high-risk groups (Q1–Q2), whereas hsa-miR-363-3p, DUSP15, and GSTM4 were upregulated in the low-risk groups (Q3–Q4), consistent with the directions revealed by the above analysis. When patients were instead stratified by the median predicted risk (Figure 3e), the high-risk group again exhibited significantly worse survival (p=1.42×10^{-3}), with AKR1E2, ARMC1, SDHAF4, TMEM242, and PCDHB2 upregulated and NDUFV2-AS1, GSTM4, DUSP15, SUSD3, TPRG1, and RFX2 downregulated in the high-risk group, further confirming that the risk signals captured by the model were aligned with actual molecular changes. Consistently, the distribution of predicted risk scores across breast cancer molecular subtypes (Figure 3h) showed that Basal and Her2 subtypes had the highest risk, LumA and LumB subtypes had markedly lower risk, and LumA had the lowest risk, matching known biological characteristics of breast cancer subtypes [53].

To validate the effectiveness of these 14 significant genes, we built a multivariable CoxPH model on the training set using their expression levels and evaluated its performance on the test set. Figure 3g illustrated that CoxPH significantly separated high-risk and low risk patients (HR=3.60, p=0.0149, C-Index=0.6998), indicating good generalization of the survival-related genes identified by LaCONIC. To further explore potential regulatory relationships, we constructed a gene interaction network based on a known ceRNA network (Figure 3f), in which hsa-miR-148a-3p, hsa-miR-363-3p, hsa-miR-365a-3p, TMEM242 and GSTM4 formed a core interaction cluster connected to risk genes such as ARMC1 and TMEM242. This finding suggested synergistic effects in tumor progression and prognosis. Consistently, multi-gene survival analyses in Figure 3i showed that several gene combinations (e.g., hsa-miR-148a-3p + hsa-miR-363-3p + GSTM4 + TMEM242, TMEM242 + ARMC1 + hsa-miR-365a-3p) significantly stratified patients into high-risk and low-risk groups (p¡0.05), with markedly poorer survival in the high-expression groups, demonstrating that the core genes identified by LaCONIC had independent prognostic value and exhibited synergistic risk effects, thereby supporting the biological plausibility and clinical relevance of our model's predictions.

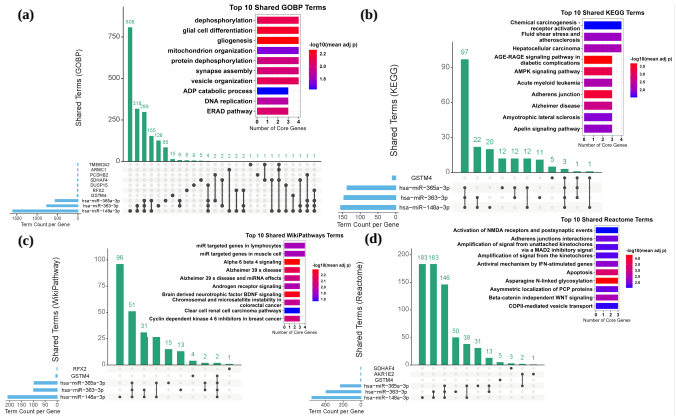

Fig. 4. Enrichment analysis of genes significantly related to survival.

2) Enrichment analysis: To investigate the biological functions of these 14 genes identified by LaCONIC, we performed pathway enrichment analysis across Gene Ontology Biological Process (GOBP), KEGG, WikiPathways and Reactome. At the GOBP level (Figure 4a), these genes were primarily enriched in dephosphorylation, glial cell differentiation, mitochondrion organization, synapse assembly, and DNA replication, implicating tumor metabolism, neuroendocrine-like transdifferentiation, and cell-cycle activity, with TMEM242, ARMC1, PCDHB2, and SDHAF4 recurrent across multiple terms [54], [55]. KEGG analysis (Figure 4b) indicated that GSTM4, hsa-miR-365a-3p, hsa-miR-363-3p, and hsa-miR-148a-3p were mainly involved in chemical carcinogenesis, AMPK signaling, and AGE–RAGE signaling, suggesting that LaCONIC-identified genes may affect survival via energy metabolism and signal transduction [56]. WikiPathways enrichment (Figure 4c) further highlighted miRNA-target regulation pathways, BDNF signaling, and the 'CDK4/6 inhibitors in breast cancer' pathway, indicating potential roles in tumor signaling, proliferation, and cell-cycle control via the CDK4/6–Cyclin D–Rb axis [57]. Reactome analysis (Figure 4d) supported

these findings by revealing enrichment in NMDA receptor signaling, adherens junctions, Wnt signaling, and apoptosis, with genes such as RFX2 and GSTM4 shared across multiple modules [58]. Together, these results showed that LaCONIC-identified genes were tightly linked to tumor-related biological processes, energy metabolism, and neural-associated regulatory networks, reinforcing the biological plausibility of the model's risk predictions.

E. Latent space visualization analysis

To further characterize the latent structure learned by LaCONIC for survival risk and molecular phenotypes, we performed UMAP-based visualization and correlation analysis on the BRCA test set. As shown in Figure 5(a), the survival-related latent vector Z_{surv} clearly separated the four risk quartiles (Q1–Q4): high-risk patients (Q1, Q2) clustered at one end of the latent space, forming a 'poor prognosis' region, whereas low-risk patients (Q3, Q4) were more dispersed and formed several subclusters, consistent with a continuous risk gradient and an underlying risk evolution trajectory. The middle and right panels showed that both Z_{surv} and Z_{cls} separated breast cancer subtypes, with sharper clusters for Z_{cls}. Jointly examining the left and middle panels further revealed that LumA samples concentrated in low-risk regions, whereas Basal samples were enriched in high-risk regions, indicating that LaCONIC aligned survival risk with subtype information and encoded both continuous risk gradients and subtype-specific molecular patterns in its latent space.

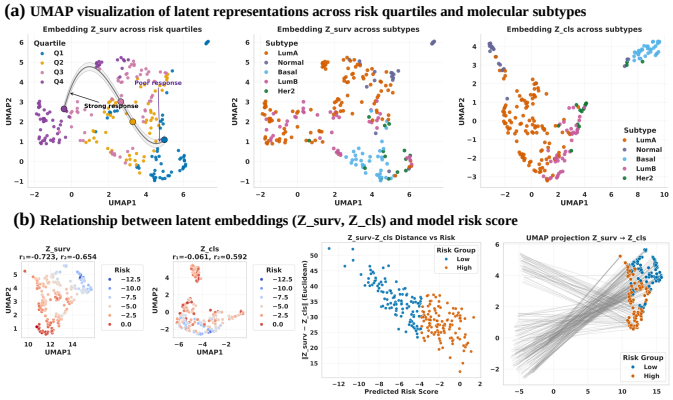

Fig. 5. Integrated survival and subtype related latent representations reveal risk-associated structure in the embedding space.

Figure 5(b) summarized the relationship between the survival representation Z_{surv}, the classification representation Z_{cls}, and the predicted risk scores. The left panels showed a strong negative correlation between Z_{surv} and risk (r = -0.723), indicating that latent position encoded a clear survival risk gradient, whereas Z_{cls} showed only weak correlation (r = -0.061). The middle panel showed that the Euclidean distance $\|Z_{\text{surv}} - Z_{\text{cls}}\|$ decreased with increasing risk, implying more compact and consistent latent representations for high-risk patients. The right UMAP projection further indicated that low-risk samples occupied a broader region between Z_{surv} and Z_{cls}, while high-risk samples converged, suggesting a tighter fusion of survival and subtype information. Together, these results indicated that LaCONIC effectively integrated survival signals and molecular features at the representation level.

F. Ablation study

To comprehensively evaluate the effectiveness of the LaCONIC model, we conducted ablation experiments on all cancer datasets from the following three aspects, and evaluated the performance using Harrell's C-Index.

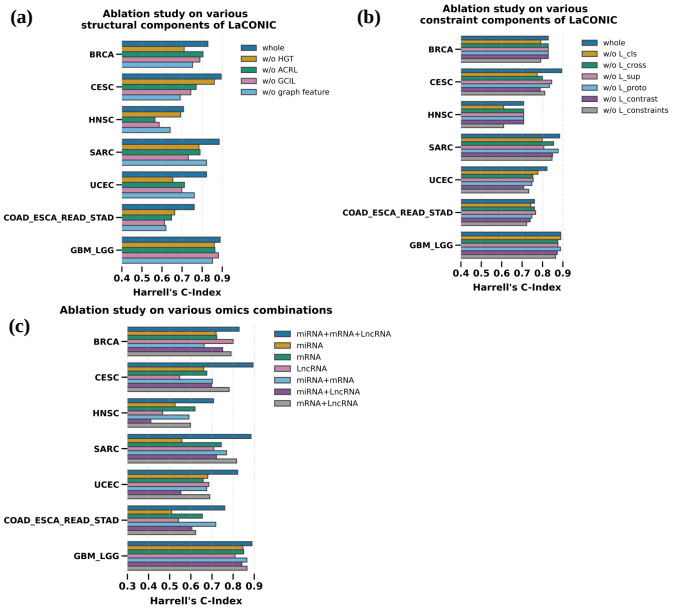

Fig. 6. Ablation experiment results based on Harrell's C-Index from different perspectives.

1) Architectural components: To assess the contribution of each architectural component in LaCONIC, we conducted ablation experiments by removing heterogeneous graph self-supervised pretraining (w/o HGT), ACRL (w/o ACRL), GCIL (w/o GCIL), and gene graph features (w/o graph feature). In the w/o HGT setting, pretrained graph embeddings were replaced by random embeddings to retain gene-level features in both modal; in the w/o graph feature setting, only gene expression features were used. As shown in Figure 6(a), the whole LaCONIC achieved the best performance across all datasets, with particularly large gains on CESC and SARC. Removing HGT or graph features caused substantial performance drops, especially on BRCA and UCEC, indicating the critical role of structured graph representations and the detrimental effect of random graph features. Eliminating ACRL or GCIL also degraded performance on most datasets, suggesting that these modules effectively capture multi-omics representations and cross-omics dependencies. Overall, LaCONIC's performance depends on the synergy of these modules.

2) Label-aware constraint: To assess the effect of the label-aware constraint mechanism, we ablated subtype classification loss (w/o L_{cls}), cross-task alignment loss (w/o L_{cross}), supervised contrastive loss (w/o L_{sup}), prototype contrastive loss (w/o L_{proto}), the combined contrastive term ($L_{\text{contrast}} = L_{\text{cross}} + L_{\text{sup}} + L_{\text{proto}}$), and the full constraint block ($L_{\text{constraints}} =$

$L_{\text{cls}}+L_{\text{contrast}}$). As shown in Figure 6(b), the complete label-aware constraint consistently yielded the best performance, and removing $L_{\text{constraints}}$ caused large drops, especially on COAD_ESCA_READ_STAD and GBM_LGG. L_{cls} was critical on most datasets and was the only non-survival loss with a strong impact on BRCA and HNSC. The impact of individual contrastive losses was cancer-specific, while removing the combined L_{contrast} produced larger performance degradations than ablating any single contrastive term, underscoring the importance of multi-level contrastive constraints for LaCONIC's generalization and survival prediction accuracy.

3) Multi-omics combinations: To quantify the contribution of different omics, we performed multi-omics ablations using miRNA, mRNA, and lncRNA individually, in all pairwise combinations, and in the full three-omics setting. As shown in Figure 6(c), integrating miRNA+mRNA+lncRNA consistently improved performance across cancers, with marked gains on CESC and SARC, while pairwise combinations generally outperformed single-omics inputs. Among single modalities, mRNA showed the most robust performance, miRNA was particularly informative in nervous system tumors (e.g., GBM_LGG), and lncRNA added value in cancers such as BRCA. To avoid hyperparameter bias, we further retuned LaCONIC on the CESC dataset for each omics combination; as reported in Supplementary Table S3, the miRNA+mRNA+lncRNA setting achieved the best results on all metrics (Harrell's C-Index = 0.8943, iAUC = 0.9093, IPCW C-Index = 0.8980), with mRNA alone ranking second and miRNA alone yielding the highest IPCW C-Index. Although lncRNA alone performed worst, its combination with mRNA produced clear gains. Overall, these results indicated complementary and synergistic effects among miRNA, mRNA, and lncRNA and showed that multi-omics integration substantially improved the accuracy and robustness of survival prediction.

V. Conclusion

Accurate cancer survival prediction is essential for guiding treatment and reducing mortality. To overcome limitations in multi-omics integration and network modeling, we proposed LaCONIC integrating heterogeneous graph self-supervised pretraining, adaptive cross-omics representation learning, graph-guided cross-omics interaction modules, and multi-level label-aware constraints to collaboratively capture molecular network topology and subtype-discriminative features. Experiments on multiple cancer datasets showed that LaCONIC consistently outperforms existing survival models, demonstrating strong representational and generalization capabilities. Interpretability analysis on breast cancer highlighted survival-associated genes that were enriched in pathways related to signal transduction, energy metabolism, cell-cycle regulation, and the "CDK4/6 inhibitors in breast cancer", supporting their mechanistic relevance. Latent space visualizations further revealed a smooth risk gradient and clearly separated molecular subtypes, underscoring the interpretability and biological consistency of the learned representations.

While LaCONIC performs well in survival prediction and interpretability, it has several limitations. First, model training relies on high-quality multi-omics paired data, so missing modalities and batch effects may hinder generalization. Second, although interpretability analyses provide some feature interpretation, experimental validation of the identified regulators and pathways is still required. In future work, we plan to explore cross-cancer transfer learning and dynamic graph modeling to capture temporal biological processes such as tumor evolution and treatment response, and to integrate single-cell omics and spatial transcriptomics data to more comprehensively characterize cellular heterogeneity and the tumor microenvironment.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (#62372165 and #62032007).

References

[1] Rebecca L Siegel, Angela N Giaquinto, and Ahmedin Jemal. Cancer statistics, 2024. *CA: a cancer journal for clinicians*, 74(1), 2024.

[2] Freddie Bray, Mathieu Laversanne, Hyuna Sung, Jacques Ferlay, Rebecca L Siegel, Isabelle Soerjomataram, and Ahmedin Jemal. Global cancer statistics 2022: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, 74(3):229–263, 2024.

[3] Douglas Hanahan. Hallmarks of cancer: new dimensions. *Cancer discovery*, 12(1):31–46, 2022.

[4] David R Cox. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202, 1972.

[5] Robert Tibshirani. The lasso method for variable selection in the cox model. *Statistics in medicine*, 16(4):385–395, 1997.

[6] Noah Simon, Jerome H Friedman, Trevor Hastie, and Rob Tibshirani. Regularization paths for cox's proportional hazards model via coordinate descent. *Journal of statistical software*, 39:1–13, 2011.

[7] Jared L Katzman, Uri Shaham, Alexander Cloninger, Jonathan Bates, Tingting Jiang, and Yuval Kluger. Deepsurv: personalized treatment recommender system using a cox proportional hazards deep neural network. *BMC medical research methodology*, 18(1):24, 2018.

[8] Changhee Lee, William Zame, Jinsung Yoon, and Mihaela Van Der Schaar. Deephit: A deep learning approach to survival analysis with competing risks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

[9] Zifeng Wang and Jimeng Sun. Survtrace: Transformers for survival analysis with competing events. In *Proceedings of the 13th ACM international conference on bioinformatics, computational biology and health informatics*, pages 1–9, 2022.

[10] Jing Teng, Lan Yang, Shan Wang, and Jing Yu. A semi-supervised transformer survival prediction model for lung cancer. *Advanced Functional Materials*, page 2419005, 2025.

[11] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[12] Eleni Anastasiadou, Leni S Jacob, and Frank J Slack. Non-coding rna networks in cancer. *Nature Reviews Cancer*, 18(1):5–18, 2018.

[13] Pan-cancer analysis of whole genomes. *Nature*, 578(7793):82–93, 2020.

[14] Marie-Anne Goyette, Marla Lipsyc-Sharf, and Kornelia Polyak. Clinical and translational relevance of intratumor heterogeneity. *Trends in cancer*, 9(9):726–737, 2023.

[15] Gustavo Arango-Argoty, Elly Kipkogei, Ross Stewart, Gerald J Sun, Arijit Patra, Ioannis Kagiampakis, and Etai Jacob. Pretrained transformers applied to clinical studies improve predictions of treatment efficacy and associated biomarkers. *Nature Communications*, 16(1):2101, 2025.

[16] Julius Keyl, Philipp Keyl, Grégoire Montavon, René Hosch, Alexander Brehmer, Liliana Mochmann, Philipp Jurmeister, Gabriel Dernbach, Moon Kim, Sven Koitka, et al. Decoding pan-cancer treatment outcomes using multimodal real-world data and explainable artificial intelligence. *Nature Cancer*, 6(2):307–322, 2025.

[17] 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.

[18] Junjun Zhang, Rosita Bajari, Dusan Andric, Francois Gerthoffert, Alexandru Lepsa, Hardeep Nahal-Bose, Lincoln D Stein, and Vincent Ferretti. The international cancer genome consortium data portal. *Nature biotechnology*, 37(4):367–369, 2019.

[19] Hua Chai, Xiang Zhou, Zhongyue Zhang, Jiahua Rao, Huiying Zhao, and Yuedong Yang. Integrating multi-omics data through deep learning for accurate cancer prognosis prediction. *Computers in biology and medicine*, 134:104481, 2021.

[20] Ruiqing Li, Xingqi Wu, Ao Li, and Minghui Wang. Hfbsurv: hierarchical multimodal fusion with factorized bilinear models for cancer survival prediction. *Bioinformatics*, 38(9):2587–2594, 2022.

[21] Xingqi Wu, Yi Shi, Minghui Wang, and Ao Li. Camr: cross-aligned multimodal representation learning for cancer survival prediction. *Bioinformatics*, 39(1):btad025, 2023.

[22] Zhimin Li, Wenlan Chen, Hai Zhong, and Cheng Liang. Pclsurv: a prototypical contrastive learning-based multi-omics data integration model for cancer survival prediction. *Briefings in Bioinformatics*, 26(2):bbaf124, 2025.

[23] Rituparno Sen, Suman Ghosal, Shaoli Das, Subrata Balti, and Jayprokas Chakrabarti. Competing endogenous rna: the key to posttranscriptional regulation. *The Scientific World Journal*, 2014(1):896206, 2014.

[24] Xiaolong Qi, Da-Hong Zhang, Nan Wu, Jun-Hua Xiao, Xiang Wang, and Wang Ma. cerna in cancer: possible functions and clinical implications. *Journal of medical genetics*, 52(10):710–718, 2015.

[25] Pei Liu, Ying Liu, Jiawei Luo, and Yue Li. Mirgraph: A hybrid deep learning approach to identify microrna-target interactions by integrating heterogeneous regulatory network and genomic sequences. In *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 1028–1035. IEEE, 2024.

[26] Yifei Yang, Shiqi Zhang, and Li Guo. Characterization of cell cycle-related competing endogenous rnas using robust rank aggregation as prognostic biomarker in lung adenocarcinoma. *Frontiers in Oncology*, 12:807367, 2022.

[27] Selcen Ari Yuka and Alper Yilmaz. Decoding dynamic mirna: cerna interactions unveils therapeutic insights and targets across predominant cancer landscapes. *Biodata Mining*, 17(1):11, 2024.

[28] Pei Liu, Xiao Liang, Yue Li, and Jiawei Luo. Convntc: convolutional neural tensor completion for detecting "a-a-b" type biological triplets. *Briefings in Bioinformatics*, 26(4):bbaf372, 2025.

[29] Hemant Ishwaran, Udaya B Kogalur, Eugene H Blackstone, and Michael S Lauer. Random survival forests. 2008.

[30] Shi Hu, Egill Fridgeirsson, Guido van Wingen, and Max Welling. Transformer-based deep survival analysis. In *Survival Prediction-Algorithms, Challenges and Applications*, pages 132–148. PMLR, 2021.

[31] Travers Ching, Xun Zhu, and Lana X Garmire. Cox-nnet: an artificial neural network method for prognosis prediction of high-throughput omics data. *PLoS computational biology*, 14(4):e1006076, 2018.

[32] Zhilong Lv, Yuexiao Lin, Rui Yan, Ying Wang, and Fa Zhang. Transsurv: transformer-based survival analysis model integrating histopathological images and genomic data for colorectal cancer. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 20(6):3411–3420, 2022.

[33] Gang Wen and Limin Li. Fgcnsurv: dually fused graph convolutional network for multi-omics survival prediction. *Bioinformatics*, 39(8):btad472, 2023.

[34] Hokeun Sun, Wei Lin, Rui Feng, and Hongzhe Li. Network-regularized high-dimensional cox regression for analysis of genomic data. *Statistica Sinica*, 24(3):1433, 2014.

[35] Wei Liu, Wei Wang, Guohua Tian, Wenming Xie, Li Lei, Jiujin Liu, Wanxun Huang, Liyan Xu, and Enmin Li. Topologically inferring pathway activity for precise survival outcome prediction: breast cancer as a case. *Molecular bioSystems*, 13(3):537–548, 2017.

[36] Wei Lan, Haibo Liao, Qingfeng Chen, Lingzhi Zhu, Yi Pan, and Yi-Ping Phoebe Chen. Deepkegg: a multi-omics data integration framework with biological insights for cancer recurrence prediction and biomarker discovery. *Briefings in bioinformatics*, 25(3):bbae185, 2024.

[37] Romana T Pop, Ping-Han Hsieh, Tatiana Belova, Anthony Mathelier, and Marieke L Kuijjer. Gene regulatory network integration with multi-omics data enhances survival predictions in cancer. *Briefings in Bioinformatics*, 26(4):bbaf315, 2025.

[38] Wenliang Zhu, Yilei Zhao, Yingqi Xu, Yong Sun, Zhe Wang, Wei Yuan, and Zhimin Du. Dissection of protein interactomics highlights microrna synergy. *PLoS One*, 8(5):e63342, 2013.

[39] Pei Liu, Jiawei Luo, and Xiangtao Chen. mircom: tensor completion integrating multi-view information to deduce the potential disease-related mirna-mirna pairs. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 19(3):1747–1759, 2020.

[40] 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.

[41] Jianjiong Gao, Bülent Arman Aksoy, Ugur Dogrusoz, Gideon Dresdner, Benjamin Gross, S Onur Sumer, Yichao Sun, Anders Jacobsen, Rileen Sinha, Erik Larsson, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cbioportal. *Science signaling*, 6(269):pl1–pl1, 2013.

[42] Mary J Goldman, Brian Craft, Mim Hastie, Kristupas Repečka, Fran McDade, Akhil Kamath, Ayan Banerjee, Yunhai Luo, Dave Rogers, Angela N Brooks, et al. Visualizing and interpreting cancer genomics data via the xena platform. *Nature biotechnology*, 38(6):675–678, 2020.

[43] Ziniu Hu, Yuxiao Dong, Kuansan Wang, and Yizhou Sun. Heterogeneous graph transformer. In *Proceedings of the web conference 2020*, pages 2704–2710, 2020.

[44] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PmLR, 2020.

[45] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017.

[46] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.

[47] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. *Advances in neural information processing systems*, 33:18661–18673, 2020.

[48] Daniel Morales Brotons, Thijs Vogels, and Hadrien Hendrikx. Exponential moving average of weights in deep learning: Dynamics and benefits. *Transactions on Machine Learning Research Journal*, 2024.

[49] Junnan Li, Pan Zhou, Caiming Xiong, and Steven Hoi. Prototypical contrastive learning of unsupervised representations. In *International Conference on Learning Representations*, 2021.

[50] Scott M Lundberg, Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su-In Lee. From local explanations to global understanding with explainable ai for trees. *Nature machine intelligence*, 2(1):56–67, 2020.

[51] Veronika Weyer and Harald Binder. A weighting approach for judging the effect of patient strata on high-dimensional risk prediction signatures. *BMC bioinformatics*, 16(1):294, 2015.

[52] Cameron Davidson-Pilon. lifelines: survival analysis in python. *Journal of Open Source Software*, 4(40):1317, 2019.

[53] Charles M Perou, Therese Sørlie, Michael B Eisen, Matt Van De Rijn, Stefanie S Jeffrey, Christian A Rees, Jonathan R Pollack, Douglas T Ross, Hilde Johnsen, Lars A Akslen, et al. Molecular portraits of human breast tumours. *nature*, 406(6797):747–752, 2000.

[54] Douglas C Wallace. Mitochondria and cancer. *Nature Reviews Cancer*, 12(10):685–698, 2012.

[55] Morgane Macheret and Thanos D Halazonetis. Dna replication stress as a hallmark of cancer. *Annual Review of Pathology: Mechanisms of Disease*, 10(1):425–448, 2015.

[56] Xiaojie Ma, Lan Jin, Xiaoqin Lei, Jingan Tong, and Runsheng Wang. Microrna-363-3p inhibits cell proliferation and induces apoptosis in retinoblastoma cells via the akt/mtor signaling pathway by targeting pik3ca. *Oncology reports*, 43(5):1365–1374, 2020.

[57] Filipa Lynce, Ayesha N Shajahan-Haq, and Sandra M Swain. Cdk4/6 inhibitors in breast cancer therapy: current practice and future opportunities. *Pharmacology & therapeutics*, 191:65–73, 2018.

[58] W Stephen Kistler, Dominique Baas, Sylvain Lemeille, Marie Paschaki, Queralt Seguin-Estevez, Emmanuèle Barras, Wenli Ma, Jean-Luc Duteyrat, Laurette Morlé, Bénédicte Durand, et al. Rfx2 is a major transcriptional regulator of spermiogenesis. *PLoS genetics*, 11(7):e1005368, 2015.