Supplementary Materials for MTPrior: A Multi-Task Hierarchical Graph Embedding Framework for Prioritizing Hepatocellular Carcinoma-Associated Genes and Long Noncoding RNAs

Fatemeh Keikha[1] and Zhi-Ping Liu[1, 2, *]

[1]Department of Biomedical Engineering, School of Control Science and Engineering, Shandong University, Jinan, Shandong 250061, China
[2]National Center for Applied Mathematics, Shandong University, Jinan, Shandong 250100, China.
[*]E-mail:zpliu@sdu.edu.cn

S1. Expanded Background

As outlined in the main manuscript, computational methodologies for identifying and prioritizing disease–gene and lncRNA associations can broadly be categorized into three distinct groups: machine learning-driven methods, network-based approaches, and graph representation learning techniques.

Initial research endeavors harnessed machine learning techniques to discern and prioritize gene-disease associations by extracting salient features from intricate networks [1]. Tran et al. [2] integrated multiple biological data sources into a cohesive network and applied a regularized SVM to predict disease-related genes, while Xu et al. [3], on the other hand, employed a random forest classifier specifically for Alzheimer's disease. The exploration of lncRNA-disease associations through computational models initiated with Chen et al.'s semi-supervised framework (LRLSLDA) [4], which was subsequently succeeded by models such as IPCARF, ILDMSF, and MFLDA. The latter models emphasized the integration of similarities and the reconstruction of association matrices [5, 6, 7]. To transcend MFLDA's constraints, Wang et al. introduced WMFLDA [8] and SelMFDF [9], while other existing models like SIMCLDA [10], and Deng et al.'s Gradient Boosting Tree, and semi-supervised frameworks [11] including LRLSLDA [4] and DNILMF-LDA [12], emerged to identifying disease-associated lncRNAs by integrating multifaceted data sources.

Numerous network-centric models have been deployed to predict disease-gene and lncRNA-disease associations, predominantly relying on random walk algorithms executed on networks grounded in the premise that functionally similar proteins are more densely interconnected within protein-protein interaction (PPI) networks [13]. These models integrate an array of data sources, encompassing gene expression profiles and protein interactions, to prioritize disease-relevant genes and anticipate lncRNA-disease linkages [14, 15]. Strategies like Vavien [13], Arete [14] and SLN-SRW [16] argument predictive capabilities by incorporating random walk with restart, and Laplacian normalization techniques. In the realm of lncRNA-disease prediction, models such as BPLLDA [17] RWRlncD [18], and GrWLDA [19] leverage random walk-based methods, whereas others like MHRWR [20], IRWRLDA [21], BRWLDA [22], and multi-layered network models including LRWRHLDA [23] integrate lncRNA and disease similarity networks with association networks. Despite their notable achievements, integrating heterogeneous similarity data to further refine prediction accuracy remains a formidable challenge.

While network-based methods demonstrate commendable performance, they confront stemming from network topology biases and intricacies in fusing multiple data sources. Matrix completion approaches, presupposing a linear gene-disease relationship, frequently struggle to encapsulate the nonlinear intricacies of reality. Recent strides in graph representation learning have bolstered gene-disease relationship by leveraging network structures [24]. Zhu et al. harnessed graph learning coupled with clustering loss enhance disease-gene predictions [25], whereas Li et al. employed GCNs on distinct networks to refine prioritization [24]. Han et al. integrated GCN with matrix factorization to grasp nonlinear relationships

[26], and the PGCN model fused genetic and disease data to amplify accuracy [24]. In the realm of lncRNA-disease prediction, models such as GAIRD [27], HGC-GAN [28], GANLDA [29], CNNDLP [30], and CapsNetLDA [31] utilize diverse techniques, including graph attention, and convolutional networks, to decipher complex patterns within lncRNA-disease networks.

Advancements in graph representation learning have ushered in fresh avenues for dissecting genomes within the biological networks [32]. These methodologies, encompassing node embedding methods techniques DeepWalk [33], and Node2Vec [34], which map nodes into vectorial representations, as well as Graph Convolutional Networks [26] and Graph Attention Networks [35] that use graph neural network architectures for node representation, adeptly encapsulate both qualitative and quantitative attributes of network nodes within a mathematically rigorous framework. Network biology harnesses these embedding models to transform intricate biological data into manageable forms, thereby facilitating similarity searches, clustering, and visualization [36]. Node embedding methods meticulously preserve the relative properties among nodes, ensuring that similar nodes are mirrored similarly in their representations, a cornerstone for effective analysis [32]. Beyond computational efficiency, there embedding techniques also capture functional properties and bolster robustness by mitigating noise [36]. As a result, graph embedding methods streamline, visualize, and enhance the analysis of complex biological networks, rendering them indispensable tools for comprehending and interpreting biological data [32, 36]. Recently, heterogeneous graph embedding has garnered significant attention for its prowess in capturing diverse structural and semantics nuances by embedding complex networks into lower-dimensional spaces. By using meta-paths, this approach delves into the various types of relationships, fostering a deeper appreciation of the diverse local structures inherent in heterogeneous graphs [37].

S2. Algorithms

Algorithm S1 The Overall Process of the Proposed Model

Input:

- The heterogeneous network $G = \{V, E, A, R\}$
- Node features $h = \{h_{(v_j)}, \forall v_j \in V\}$
- Node labels $l = \{l_{(v_j)}, \forall v_j \in V\}$
- The meta-path set $\Phi_{\text{het}} = \{\Phi_1, \ldots, \Phi_p\}$
- Maximum Meta-path Hops Depth D
- Number of attention heads M

Output:

- Final node embeddings Z
- Final node ranking R

$\Phi_{\text{het}}^{\text{New}}, d \leftarrow$ Optimal Multi-Hop Finder Module$(G, h, \Phi_{\text{het}}, D, M)$
$Z \leftarrow$ Multi-hop Embedding Module$(G, h, \Phi_{\text{het}}^{\text{New}}, d, M)$
MLP$(Z, \Phi_{\text{het}}^{\text{New}}, l)$, Predict $f^* = \{f^*_{(v_j)}, \forall v_j \in V\}$
$R \leftarrow$ Assign ranks in descending order of f^*
return Z, R

Algorithm S2 Optimal Multi-hop Finder Module

Input:

- The heterogeneous network G
- Node features h
- Meta-path set Φ_{het}
- Maximum Meta-path Hops Depth D
- Number of attention heads M
- Node labels l

Output:

- New multi-hop meta-path space $\Phi_{\text{het}}^{\text{New}} = \{\Phi^{(\text{hop}_1)}, \ldots, \Phi^{(\text{hop}_d)}\}$
- Depth d

$\Phi^{(\text{hop}_1)} \leftarrow \Phi_{\text{het}}$
$\Phi_{\text{set}} \leftarrow \Phi^{(\text{hop}_1)}$
$\Phi_{\text{het}}^{\text{New}} \leftarrow \Phi_{\text{set}}$
$BestValue \leftarrow \text{Evaluate}(G, h, \Phi^{(\text{hop}_1)}, 1, M, l)$
for $y = 2$ **to** D **do**
    for $\Phi_i \in \Phi_{\text{set}}, i \in \{1, \ldots, P\}$ **do**
        $\Phi_{\text{temp}} \leftarrow \Phi_i^{(\text{hop}_{(y-1)})} + \Phi_i^{(\text{hop}_1)}$
        $TempSubset \leftarrow \Phi_{\text{het}}^{\text{New}} \cup \Phi_{\text{temp}}$
        $Temp \leftarrow \text{Evaluate}(G, h, TempSubset, y, M, l)$
        if $Temp > BestValue$ **then**
            $\Phi^{(\text{hop}_y)} \leftarrow \Phi^{(\text{hop}_y)} + \Phi_{\text{temp}}$
            $BestValue \leftarrow Temp$
            $d \leftarrow y$
        end if
    end for
end for
    return $\Phi_{\text{het}}^{\text{New}}, d$

Algorithm S3 Multi-hop Embedding Module

Input:

- The heterogeneous network G, Node features h
- New meta-path set $\Phi_{\text{het}}^{\text{New}}$, Maximum Meta-path Hops Depth D
- Number of attention heads M

Output:

- The final node embedding Z

for $\Phi_i^{(\text{hop}_1)} \in \{\Phi_{\text{het}}^{\text{New}}\}, i \in \{1, \ldots, q\}$ **do**
  for $\Phi_i^{(\text{hop}_y)} \in \{\Phi_i^{(\text{hop}_1)}, \ldots, \Phi_i^{(\text{hop}_d)}\}, y \leq d$ **do**
    for $m = 1 \ldots M$ **do**
      for $v_j \in V$ **do**
        Find meta-path-based neighbors $\Phi_i^{(\text{hop}_y)}$
        for $v_k \in N_{(v_j)}^{(\Phi_i^{(\text{hop}_y)})}$ **do**
          Calculate the weight coefficient $\alpha_{(v_j v_k)}^{(\Phi_i^{(\text{hop}_y)})}$
        end for
        Learn node-level embedding:

$$z_{(v_j)}^{(\Phi_i^{(\text{hop}_y)})} \leftarrow \sigma\left(\sum_{v_k \in N_{(v_j)}^{(\Phi_i^{(\text{hop}_y)})}} \alpha_{(v_j v_k)}^{(\Phi_i^{(\text{hop}_y)})} \cdot h_{(v_k)}\right)$$

      end for
    end for
    Concatenate learned embeddings from all attention heads:

$$z_{(v_j)}^{(\Phi_i^{(\text{hop}_y)})} \leftarrow \Big\|_{m=1}^{M} \sigma\left(\sum_{v_k \in N_{(v_j)}^{(\Phi_i^{(\text{hop}_y)})}} \alpha_{(v_j v_k)}^{(\Phi_i^{(\text{hop}_y)})} \cdot h_{(v_k)}\right)$$

    Calculate weight $\beta^{(\Phi_i^{(\text{hop}_y)})}$ of information-flow $\Phi_i^{(\text{hop}_y)}$ in Φ_i
  end for
  Learn information-flow level node embedding:

$$Z^{(\Phi_i)} \leftarrow \sum_{y=1}^{d} \beta^{(\Phi_i^{(\text{hop}_y)})} \cdot z_{(v_j)}^{(\Phi_i^{(\text{hop}_y)})}$$

  Calculate the meta-path attention weight vector $\gamma^{(\Phi_i)}$
end for
Fuse the meta-path embeddings: $\texttt{Z} = \sum_{i=1}^{q} \gamma^{(\Phi_i)} \cdot Z^{(\Phi_i)}$
Calculate Cross-Entropy loss function: $L = -\sum_{j=1}^{N} Y_j \cdot \log(C \cdot Z_j)$
Backpropagation and update model parameters
return Z

S3. Details of highly related methods

- Vavien [13] is a network-based method that leverages the topology of protein-protein interaction networks and employs random walk with restart to prioritize candidate disease genes based on their topological similarity to known disease genes.
- Arete [14] is another network-based method that combines biological network structure with

various types of evidence to prioritize candidate genes. It offers flexible and interoperable solutions for enhancing gene prioritization beyond traditional network-based methods, also utilizing random walk with restart.

- netSVM [38] develops an integrated approach using network-constrained support vector machines to identify cancer biomarkers by integrating gene expression data with protein-protein interaction networks.
- Hetio [39] constructs a heterogeneous network and extracts features describing the network topology between genes and diseases. It trains a model on GWAS associations to predict associations between protein-coding genes and complex diseases.
- MALANI [40] integrates gene expression data across multiple cancer types employs SVM models trained on gene-wise and gene-pair interactions. It assesses thousands of genes for their potential roles in cancer networks and uncovers novel candidates associated with cancer outcomes.
- PGCN [24] involves the use of a graph convolutional network to embed a heterogeneous network of genes and diseases, along with their features, into a latent space for further analysis.
- RWRHLD [41] is a rank-based method that utilizes a random walk with restart to prioritize candidate lncRNA-disease associations. The study integrates an lncRNA-lncRNA crosstalk network based on shared miRNA response elements, disease-disease similarity, and known lncRNA-disease association networks into a heterogeneous network.
- RisklncRNAs [42] presents a method for prioritizing risk lncRNAs in cancer by constructing a Gene-LncRNA Co-expression Network, integrating expression and protein interaction data, and applying a random walk algorithm combined with disease phenotype similarity scores.
- TPGLDA [43] is a computational method that predicts lncRNA-disease associations by integrating lncRNA-disease and gene-disease associations into a tripartite graph. It effectively captures the heterogeneity of coding and noncoding gene-disease relationships.
- LRWRHLDA [23] introduces a computational framework for predicting lncRNA-disease associations by constructing isomorphic networks for lncRNA, disease, gene, and miRNA similarities, integrating them into heterogeneous networks, and applying a Laplace normalized random walk with restart algorithm to predict associations.
- GANLDA [29] integrates heterogeneous data of lncRNAs and diseases, applies PCA for noise reduction, utilizes graph attention mechanisms to extract relevant features, and employs a multilayer perceptron for association inference.
- GAIRD [27] employs a hybrid random walk strategy to integrate diverse network information. It includes a module to separate and refine attributes and structures, utilizing techniques like group convolution and deep separable convolution to enhance feature learning.
- LDAGM [44] uses a Graph Convolutional Autoencoder and Multilayer Perceptron to predict lncRNA-disease associations. It fuses similarities from six homogeneous networks into a multi-view heterogeneous network, extracts nonlinear and deep topological features, and employs an enhanced MLP with an aggregation layer for improved prediction accuracy.
- MMHGAN [45] is a deep learning model that predicts lncRNA-disease associations by leveraging hierarchical graphical attention networks. It constructs heterogeneous and homogeneous graphs, uses multihead attention for feature aggregation, and evaluates metapath importance to extract features. Predictions are made by reconstructing these features through a fully connected layer.

The ACC metric indicates that our method accurately classifies genes and lncRNAs related to the disease. The AUROC metric demonstrates its effectiveness in distinguishing between positive and negative samples, while the AUPRC metric confirms that our method predicts a high percentage of true disease-related genes and lncRNAs. The results show that our proposed method achieves the highest scores in ACC, AUROC, and AUPRC metrics for both gene and lncRNA prediction. Specifically, for gene prediction, our method achieves scores of 0.9238, 0.95686, and 0.94675, respectively. For lncRNA prediction, our method achieves scores of 0.89526, 0.93494, and 0.90184, respectively. Additionally, MALANI ranks second in ACC, AUROC, and AUPRC for gene prioritization, with scores of 0.82828,

0.86471, and 0.88534, respectively. For lncRNA prioritization models, GANLDA ranks second in ACC with a score of 0.88666, while RWRHLD achieves the second-highest scores in AUROC and AUPRC, with values of 0.88547 and 0.88356, respectively.

The Matthews Correlation Coefficient (MCC) metric demonstrates that our method effectively aligns predictions with true disease-related genes and lncRNAs. The F1 score further confirms its high precision and recall in predicting these targets. Specifically, our method ranks first in MCC and F1 for gene prioritization, with scores of 0.84761 and 0.92380, respectively. Hetio follows in second place, with scores of 0.7912 and 0.80061, respectively, in these metrics. For lncRNA prediction, our proposed method ranks third in MCC, with a score of 0.76969, while GANLDA and LRWRHLDA secure the first and second ranks with scores of 0.83333 and 0.81818, respectively. However, in the same comparison category, our proposed model achieves the highest score of 0.86484 in F1, ranking first, GAIRD ranks second with a score of 0.84523.

Sensitivity reflects the effectiveness of our method in detecting true disease-related genes and lncRNAs, while specificity demonstrates its proficiency in correctly identifying non-disease-related cases. Our method excels in both gene and lncRNA prioritization models, achieving the highest scores in sensitivity and specificity metrics. Specifically, for gene prediction, our method scores 0.928844 and 0.940369 in sensitivity and specificity, respectively, outperforming Hetio, which ranks second with scores of 0.869047 and 0.870012. In the realm of lncRNA prediction, our model similarity shines, scoring 0.903771 and 0.903816 in sensitivity and specificity, respectively, earning the top spot among models predicting lncRNA-disease associations. LRWRHLDA follows closely in the second place with scores of 0.90123 and 0.901452. Sensitivity underscores the effectiveness of our method in detecting true disease-related genes and lncRNAs, while specificity underscores its proficiency in accurately identifying non-disease-related cases. The results clearly demonstrate that our proposed method holds an advantage over other comparison approaches in predicting disease-related genes and lncRNAs.

The proposed method attains an average AUROC of 0.9459, making a substantial 24.72% improvement over network-based methods that average an AUROC of 0.75841. These network-based methods employ the random walk with restart algorithm on biological network topologies to prioritize disease-related genes or lncRNAs. When compared to ML-based models trained on integrated gene expression data and protein-protein interaction networks, our method also demonstrates a notable performance enhancement of 23.99% in average AUROC. Furthermore, our method surpasses GNN-based models by 10.98%, underscoring the significant impact of integrating heterogeneous data, attributes, and structures with embedding models on enhancing prediction performance.

We attribute the enhanced performance of our method to several pivotal factors. We refine the backbone of the heterogeneous network to facilitate the identification of more pertinent meta-paths associated with the disease, acknowledging that the backbone nodes encompass those not directly targeted in our prediction. Furthermore, beyond merely considering direct meta-path neighbors, we delineate neighbors and their interconnections through multiple iterations of each meta-path, regarding them as multi-hop meta-path neighbors. This strategy delves deeper into the network's structure. Additionally, we deploy a two-layered hierarchical attention model that captures the attention weights of various nodes, different multi-hop meta-paths, thereby augmenting the method's precision in prioritizing disease-related nodes.

S4. Experimental results with varying sample ratios

Experiments were conducted utilizing diverse ratios of positive to negative samples to investigate their influence on the proposed method for prioritizing genes and lncRNAs. Notably, variations in the ratios between positive and negative samples had a substantial impact on the method's performance.

To assess the model's capacity to classify and prioritize genes and lncRNAs using different sample ratios, multiple experimental datasets were generated. Each dataset consists all positive samples coupled with varying quantities of negative samples: an equivalent number, five times, or ten times the amount of negative samples compared to positives. These curated datasets were employed to evaluate the prediction model's performance in relation to both genes and lncRNAs associated with HCC.

Figure 1 and Table 1 present the results, demonstrating that the proposed method attains average ACC, MCC, and F1 scores of 92.38%, 84.76%, and 92.38%, respectively, for predicting disease genes, and 89.52%, 76.96%, and 86.48%, respectively, for predicting HCC-related lncRNAs, under an equivalent ratio of positive to negative samples. Notably, our method MTPrior achieves the highest Sensitivity and Specificity values with 1:1 ratio of negative to positive samples, particularly excelling in gene prioritization tasks. Additionally, MTPrior demonstrates superior performance in terms of AUROC and AUPRC metrics with a 1:1 ratio for both gene and lncRNA classification, achieving values of 0.9568 and 0.9349

for AUC, and 0.9467 and 0.9018 for AUPRC in gene and lncRNA prediction, respectively. However, for Accuracy, the highest scores are achieved with a 1:5 ratio, yielding 0.9298 for gene prioritization and 0.9025 for lncRNA prioritization. The high AUC and AUPRC scores demonstrates the effectiveness of the proposed method in accurately distinguishing between positive and negative samples. The minimal variation in results across different ratios of positive to negative samples underscores the method's stability and reliability.

Table S1: **Results of the proposed methods under varying ratios (Positive Samples : Negative Samples = 1:1, 1:5, and 1:10).**

Method	Ratio	ACC	AUROC	AUPRC	MCC	F1	Sensitivity	Specificity
Gene prioritizing	1:1	**0.9238**	**0.9569**	**0.9468**	**0.8476**	**0.9238**	**0.9288**	**0.9404**
	1:5	0.9299	0.9251	0.9353	0.7500	0.8742	0.8355	0.8410
	1:10	0.8941	0.8960	0.8940	0.8004	0.8998	0.7485	0.7483
lncRNA prioritizing	1:1	**0.8953**	**0.9349**	**0.9018**	**0.7697**	**0.8648**	**0.9038**	**0.9038**
	1:5	0.9025	0.9021	0.8266	0.6977	0.8423	0.9023	0.9023
	1:10	0.8788	0.8636	0.8829	0.7736	0.8629	0.8674	0.8694

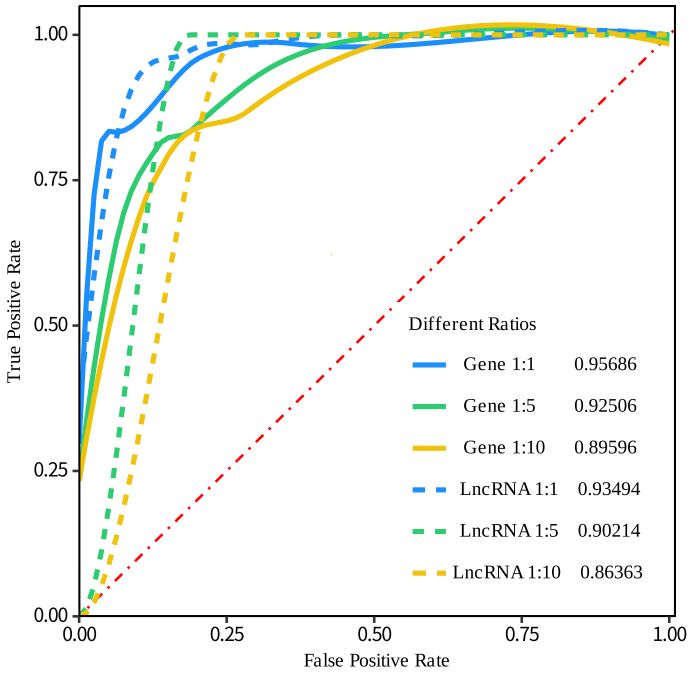


Figure S1: **AUC values for experimental results using different ratios of positive to negative samples.** Solid lines depict the gene identification task, while dashed lines represent the lncRNA iden-tification task. The blue curves indicate a 1:1 sample ratio, green curves indicate a 1:5 sample ratio, and yellow curve indicates a 1:10 sample ratio.

S5. Evaluation metrics across varying embedding sizes

We examine the influence of embedding dimension on our proposed method MTPiror, with the outcomes presented in Fig 2. The embedding sizes considered here encompass 32, 64, 128, 256, 512, and 1024, respectively. Across all evaluation metrics, our method demonstrates optimal performance in gene and lncRNA prioritization when the embedding dimensions are set at 256 and 512. Specifically, for gene prioritization, the performance of our method enhances as the embedding dimension increases from 32 to 256, while for lncRNA prioritization, it improves from 32 to 512. Nevertheless, a decline in performance is observed when the embedding dimension is escalated to 1024. Consequently, when utilizing our proposed method, we select an embedding size of 256 for gene prioritization and 512 for lncRNA prioritization.

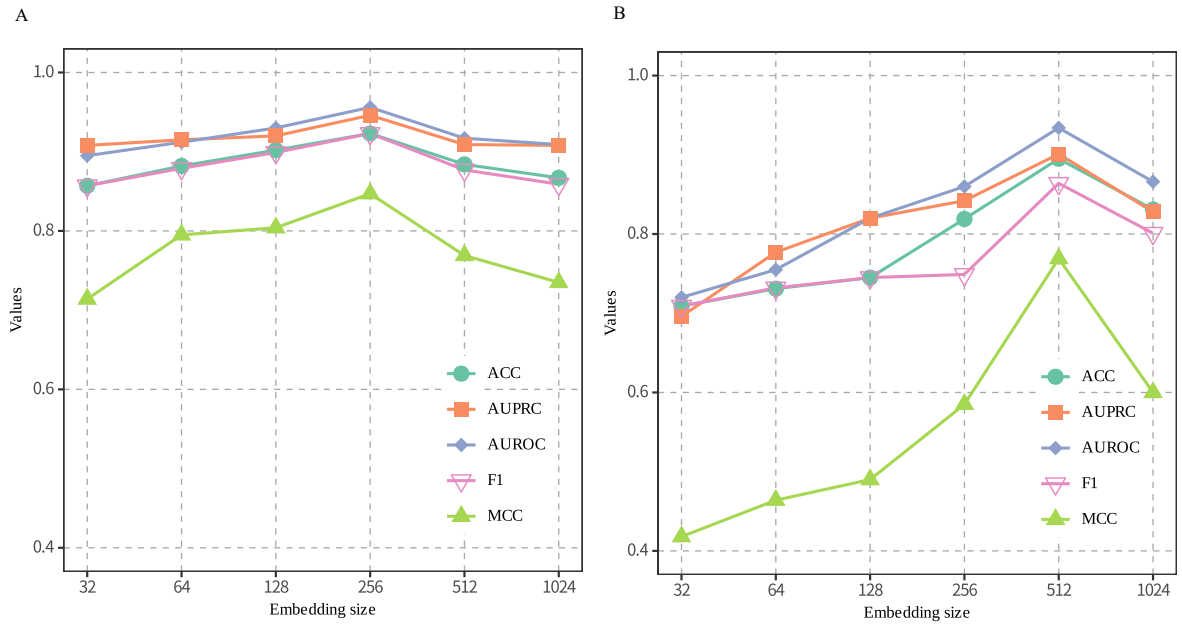


Figure S2: **Evaluation metrics across various embedding dimensions.** It contrasts the results of evaluation metrics (ACC, AUROC, AUPRC, F1, MCC) for different embedding sizes in the context of (A) gene identification and (B) lncRNA identification tasks. Each line depicts the performance of a particular evaluation metric as the embedding size varies from 32 to 1024.

S6. Evaluation metrics across different classification methods

In this experiment, we evaluated the performance of various classifiers by comparing a MLP with several other models, namely SVM, Random Forest, logistic regression, Adaboost, Naïve Bayes, and XGBoost. We employed five-fold cross-validation across multiple evaluation criteria to ensure a comprehensive assessment. To maintain fairness and reliability in our comparison, all experimental settings were kept consistent, with the exception of the classifier being tested. Our results, as illustrated in Fig 3, demonstrate that in the context of gene and lncRNA prioritization, the MLP consistently outperforms the other models across nearly all metrics. The only exception is in the AUPRC metric for the lncRNA experiment, where Random Forest achieved a marginally superior result (0.90298) compared to MLP (0.90184).

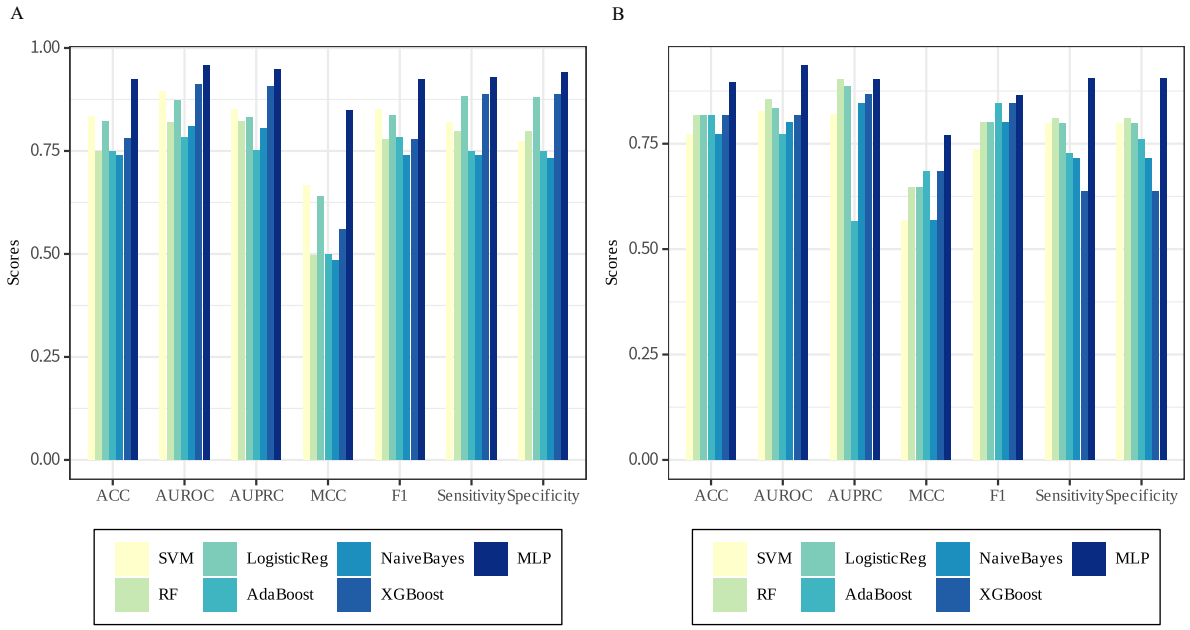


Figure S3: **Evaluation metrics across different classification methods.** It compares the different classifier's performance including SVM, Random Forest, logistic regression, Adaboost, Naïve Bayes, XGBoost, and MLP by several evaluation metrics (ACC, AUROC, AUPRC, F1, MCC) for (A) gene identification task, and (B) lncRNA identification task.

S7. Evaluation metrics in different number of attention heads

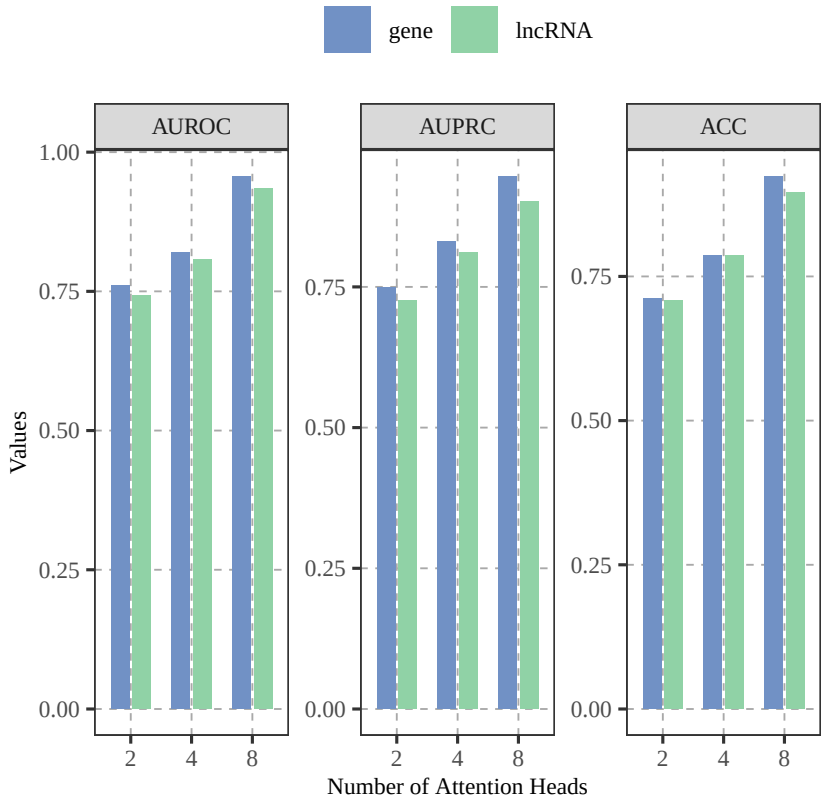


Figure S4: **Evaluation metrics in different number of attention heads.** It contrasts the results of evaluation metrics (ACC, AUROC, AUPRC) for different number of attention heads in the context of gene identification and lncRNA identification tasks.

S8. Evaluation metrics in different dropout rates

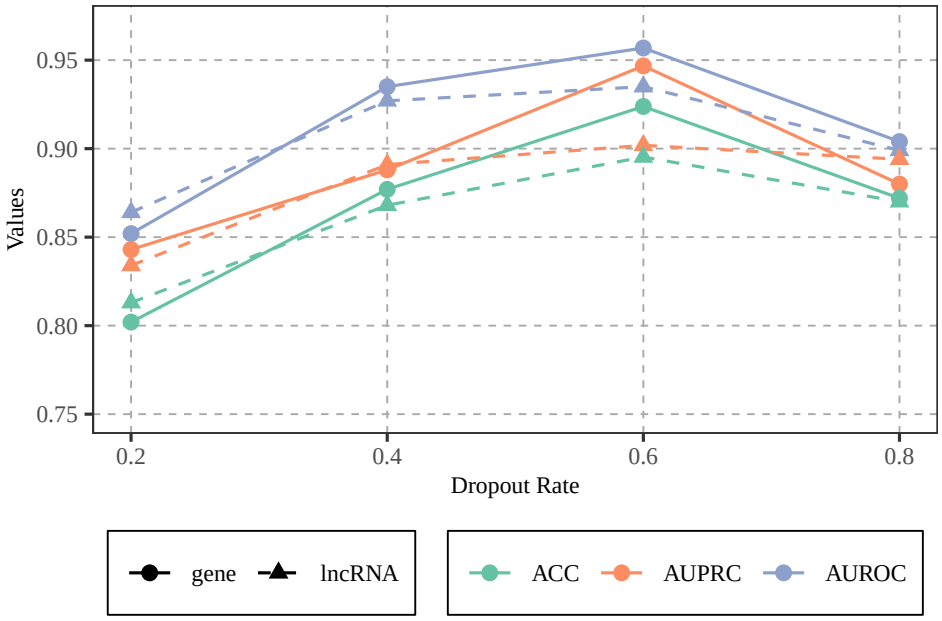


Figure S5: **Evaluation metrics in different dropout rates in the hierarchical embedding step.** It contrasts the results of evaluation metrics (ACC, AUROC, AUPRC) for different dropout rates in the context of gene identification and lncRNA identification tasks in the hierarchical embedding step. Each line depicts the performance of a particular evaluation metric as the dropout rates varies from 0.2 to 0.8.

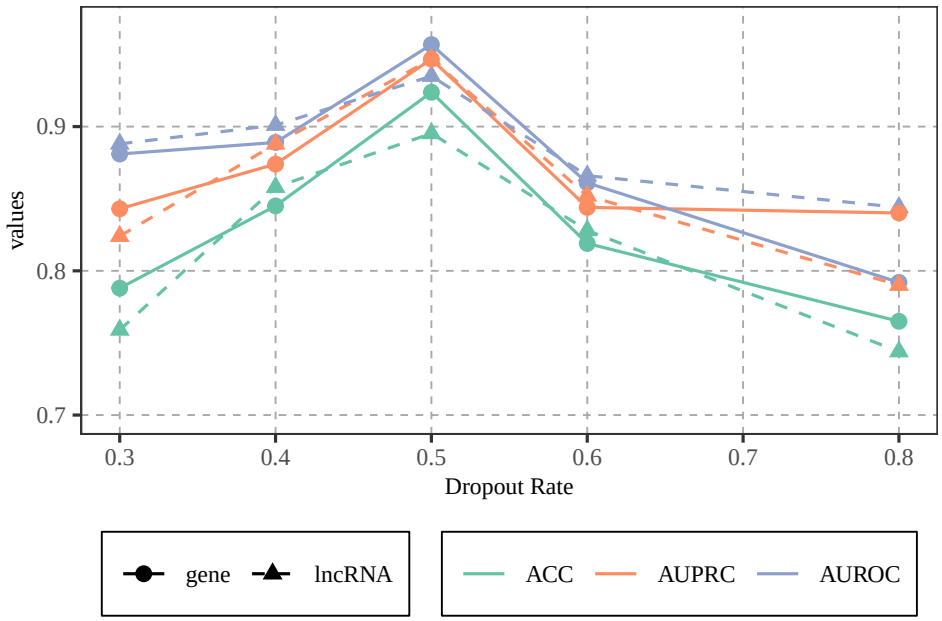


Figure S6: **Evaluation metrics in different dropout rates in the final MLP classifier.** It contrasts the results of evaluation metrics (ACC, AUROC, AUPRC) for different dropout rates in the context of gene identification and lncRNA identification tasks in the final MLP classifier. Each line depicts the performance of a particular evaluation metric as the dropout rates varies from 0.3 to 0.8.

S9. Evaluation metrics across different classification methods

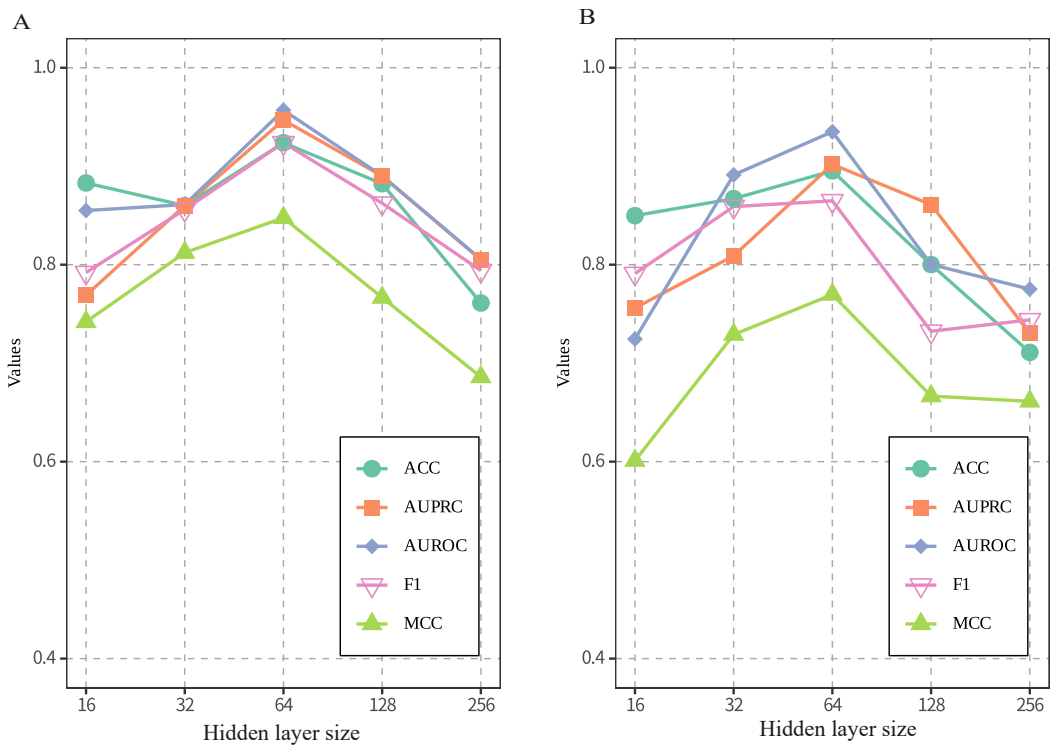


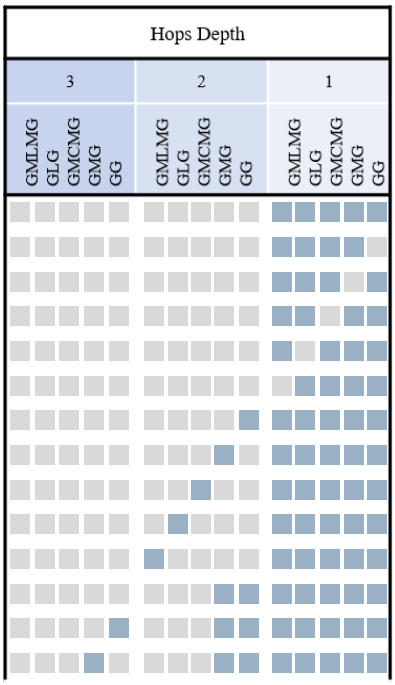
Figure S7: **Evaluation metrics across various hidden layer size.** It contrasts the results of evaluation metrics (ACC, AUROC, AUPRC, F1, MCC) for different hidden layer size in the context of (A) gene identification and (B) lncRNA identification tasks. Each line depicts the performance of a particular evaluation metric as the hidden layer size varies from 16 to 256.

S10. Time complexity of MTPrior

As explained in the *Complexity analysis* section, the time complexity of MTPrior is primarily influenced by the operations within the *Optimal Multi-hop Finder Module* and the *Multi-hop Embedding Module*. In the *Optimal Multi-hop Finder Module*, the time complexity arises from the iterative examination of meta-paths spanning multiple hops. In the *Multi-hop Embedding Module*, the time complexity is determined by the computation of node embeddings across various meta-path hops. This results in an overall time complexity for MTPrior of $\mathcal{O}(D \times P \times A) + A$, where $A = \mathcal{O}\left(D \times P\left(|V|Md + |E|Md\right) + (D \times P) \times (|V|d)\right)$. Upon rearrangement, the time complexity can be expressed as $\mathcal{O}\left((DP+1)(DPd)\left((M+1)|V| + M|E|\right)\right)$, where D represents the maximum meta-path hop depth, P denotes the number of initial meta-paths, $|V|$ and $|E|$ represent the number of nodes and meta-path-based node pairs, respectively, M is the number of attention heads, and d is the out-feature dimension.

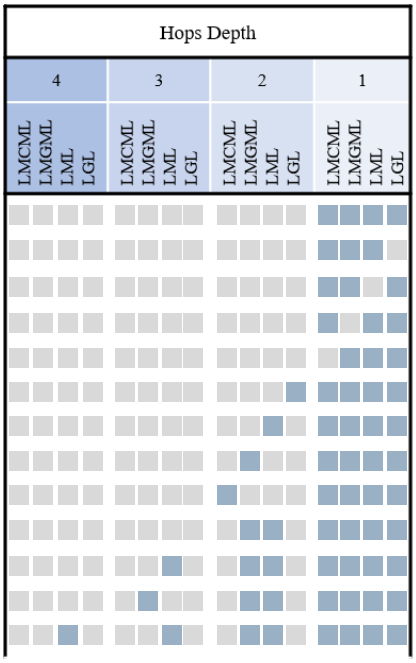
In this study, the output dimensions were set to 256 for genes and 512 for lncRNAs, the number of attention heads was set to 8, and the number of initial meta-paths was set to 5 and 4 for gene and lncRNA prioritization, respectively. Consequently, the time complexity of MTPrior is related to the number of initial meta-paths, meta-path hop depth, the number of nodes, and edges.

To provide a quantitative understanding, we computed the time complexity for each combination of multi-hop meta-paths listed in Figure 5 and presented the results in Figure S4. The results indicate that the time complexity scales linearly with the number of nodes and edges.



Gene prioritization

D	P	M	d	Time complexity $O((DP+1)(DPd)((M+1)\|V\|+M\|E\|))$
1	5	8	256	$O(69120\|V\|+61440\|E\|)$
1	4	8	256	$O(46080\|V\|+40960\|E\|)$
1	4	8	256	$O(46080\|V\|+40960\|E\|)$
1	4	8	256	$O(46080\|V\|+40960\|E\|)$
1	4	8	256	$O(46080\|V\|+40960\|E\|)$
1	4	8	256	$O(46080\|V\|+40960\|E\|)$
2	5	8	256	$O(253440\|V\|+225280\|E\|)$
2	5	8	256	$O(253440\|V\|+225280\|E\|)$
2	5	8	256	$O(253440\|V\|+225280\|E\|)$
2	5	8	256	$O(253440\|V\|+225280\|E\|)$
2	5	8	256	$O(253440\|V\|+225280\|E\|)$
2	5	8	256	$O(253440\|V\|+225280\|E\|)$
3	5	8	256	$O(552960\|V\|+491520\|E\|)$
3	5	8	256	$O(552960\|V\|+491520\|E\|)$



LncRNA prioritization

D	P	M	d	Time complexity $O((DP+1)(DPd)((M+1)\|V\|+M\|E\|))$
1	4	8	512	$O(92160\|V\|+81920\|E\|)$
1	3	8	512	$O(55296\|V\|+49152\|E\|)$
1	3	8	512	$O(55296\|V\|+49152\|E\|)$
1	3	8	512	$O(55296\|V\|+49152\|E\|)$
1	3	8	512	$O(55296\|V\|+49152\|E\|)$
2	4	8	512	$O(331776\|V\|+294912\|E\|)$
2	4	8	512	$O(331776\|V\|+294912\|E\|)$
2	4	8	512	$O(331776\|V\|+294912\|E\|)$
2	4	8	512	$O(331776\|V\|+294912\|E\|)$
2	4	8	512	$O(331776\|V\|+294912\|E\|)$
3	4	8	512	$O(718848\|V\|+638976\|E\|)$
3	4	8	512	$O(718848\|V\|+638976\|E\|)$
4	4	8	512	$O(1253376\|V\|+1114112\|E\|)$

Figure S8: **Time complexity of MTPrior,** It confirms the time complexity scales linearly with the number of nodes and edges.

S11. The top 20 predicted results of HCC-related genes and lncRNA

Table S2. The top 20 predicted results of HCC-related genes based on MTPrior with the summaries of their functions.

Gene Symbol	Gene Name	Known Functions	Evidence
CDKN2A	Cyclin Dependent Kinase Inhibitor 2A	CDKN2A is associated with multiple cancers, including melanoma, pancreatic adenocarcinoma, esophageal carcinoma, and head and neck squamous cell carcinoma. CDKN2A is highly expressed in HCC and is associated with poorer survival outcomes. It plays a key role in immune regulation by correlating with increased infiltration of immune cells, including T cells, B cells, macrophages, neutrophils, and dendritic cells, suggesting its involvement in shaping the immune microenvironment and serving as a potential prognostic biomarker in HCC [46].	PMID: 34405225
CDC20	Cell Division Cycle 20	CDC20 is associated with Hepatocellular Carcinoma, Oocyte Maturation Arrest 14, Prostate Cancer, Adenocarcinoma, and various other malignancies. CDC20 is highly expressed in hepatocellular carcinoma (HCC) and is associated with poor prognosis. It promotes HCC progression by regulating epithelial-mesenchymal transition (EMT), enhancing cell proliferation, migration, and invasion. CDC20 silencing suppresses these malignant behaviors and increases radiosensitivity, especially in p53-mutant HCC cells, by affecting the Bcl-2/Bax pathway and enhancing G2/M arrest and apoptosis [47, 48].	PMID: 33907851; 34512169
CCN2	Cellular Communication Network Factor 2	Diseases associated with CCN2 include spondyloepimetaphyseal dysplasia (Li-Shao-Li type), kyphomelic dysplasia, and various fibrotic and cardiovascular conditions. CCN2 contributes to oxaliplatin resistance and poor prognosis in HCC by activating the MAPK/Id-1 signaling loop, enhancing stemness and metastasis. It is also upregulated in fibrotic livers and promotes liver fibrosis by activating hepatic stellate cells through the Slit2/Robo-PI3K/AKT pathway [49, 50].	PMID: 36469291; 36469291

Continued on next page

Table S2 Continued from previous page

Gene Symbol	Gene Name	Known Functions	Evidence
BIRC5	Baculoviral IAP Repeat Containing 5	BIRC5 is associated with diseases such as adenocarcinoma and malignant pleural mesothelioma. It is involved in pathways related to mitotic spindle formation and gene transcription, and its functions include protein homodimerization and heterodimerization. BIRC5 is highly expressed in hepatocellular carcinoma (HCC) and is associated with poor overall survival and increased infiltration of CD8+ T cells and macrophages. Silencing BIRC5 inhibits the PPARγ pathway, suppresses malignant behaviors of HCC cells, reduces tumorigenesis in vivo, and promotes cuproptosis, suggesting its potential as a therapeutic target in HCC [51].	PMID: 39585552
FOXA3	Forkhead Box A3	FOXA3 is associated with diseases such as adult-onset severe asthma and adult acute monocytic leukemia. It plays roles in pathways related to beta-cell development and FOXA transcription factor networks, with functions involving DNA-binding and transcription factor activity. FOXA3 functions as a tumor suppressor in hepatocellular carcinoma (HCC), with its downregulation linked to increased malignancy and poor prognosis. It inhibits HCC progression by regulating metabolism, repressing metastasis-related pathways, and enhancing sensitivity to anticancer treatments [52, 53].	PMID: 35764883
NR0B1	Nuclear Receptor Subfamily 0 Group B Member 1	NR0B1 is associated with congenital adrenal hypoplasia and 46,XY sex reversal, both of which involve disorders of adrenal and sexual development. NR0B1 is highly expressed in HCC and linked to poor prognosis. It promotes proliferation, migration, and invasion of HCC cells. NR0B1 enhances sorafenib resistance by activating autophagy and inhibiting apoptosis [54, 55].	PMID: 37991109

Continued on next page

Table S2 Continued from previous page

Gene Symbol	Gene Name	Known Functions	Evidence
GBP5	Guanylate Binding Protein 5	GBP5 is a protein-coding gene linked to inflammatory and immune-related conditions, including pulpitis and Cinca syndrome. It plays a role in antiviral defense and cytokine signaling pathways. Functionally, GBP5 is involved in GTPase activity and protein binding. In HCC, GBP5 was identified as a key gene closely associated with immune infiltration and tumor microenvironment (TME) scores. Higher GBP5 expression correlated with favorable immune and stromal scores and improved survival outcomes. These findings suggest that GBP5 may serve as a potential prognostic marker and indicator of immunotherapy response in HCC [56].	PMID: 33897701
GSDME	Gasdermin E	GSDME is linked to certain medical conditions, such as autosomal dominant deafness type 5 and rare forms of genetic deafness. It is involved in key biological processes including the intrinsic apoptosis pathway and programmed cell death. GSDME plays a complex role in hepatocellular carcinoma (HCC). It mediates pyroptosis through caspase-3 activation under heat stress, but its overexpression promotes tumor growth and correlates with poor patient outcomes after radiofrequency ablation (RFA). High GSDME levels are linked to resistance to PD-L1 inhibitors by increasing exhausted T cells and driving immunosuppressive M2-like macrophage polarization via the PI3K-AKT pathway. Despite being a tumor suppressor in many cancers, GSDME's elevated expression in HCC suggests it may both promote and inhibit tumor progression, highlighting its dual role and importance as a potential therapeutic target [57, 58, 59].	PMID: 37862939

Continued on next page

Table S2 Continued from previous page

Gene Symbol	Gene Name	Known Functions	Evidence
S100A8	S100 Calcium Binding Protein A8	S100A8 gene is linked to several diseases including cystic fibrosis, peptic ulcer disease, and duodenal ulcer. It is involved in immune-related pathways such as the Toll Like Receptor 7/8 (TLR7/8) cascade and other immune system disorders. The protein encoded by S100A8 plays roles in calcium ion binding and interaction with the RAGE receptor, contributing to its involvement in inflammatory and immune responses. S100A8 plays a significant role in hepatocellular carcinoma (HCC) progression, particularly in non-inflammation-driven liver tumors. Loss of S100A8/A9 reduces tumor cell proliferation and tumor size in carcinogen-induced HCC models. Additionally, hypomethylation and overexpression of S100A8 in HCC are linked to increased tumor growth, migration, and invasion, as well as poorer patient survival, suggesting that S100A8 acts as a tumor promoter and may serve as a diagnostic and prognostic biomarker for HCC [60, 61].	PMID: 27462864; 26404752
ABCB11	ATP Binding Cassette Subfamily B Member 11	ABCB11 is linked to diseases such as Progressive Familial Intrahepatic Cholestasis type 2 and Benign Recurrent Intrahepatic Cholestasis type 2. It plays a role in biological pathways related to bile acid and bile salt metabolism, as well as the generalized pharmacokinetics of statins. ABCB11 encodes the bile salt export pump (BSEP), which transports bile salts across hepatocyte membranes in an ATP-dependent manner, maintaining bile acid and lipid homeostasis. It is downregulated in HCC, and inherited mutations are linked to early-onset cases, suggesting its potential role in liver cancer development [62, 63].	PMID: 25016225

Continued on next page

Table S2 Continued from previous page

Gene Symbol	Gene Name	Known Functions	Evidence
OSMR	Oncostatin M Receptor	Diseases associated with OSMR (Oncostatin M Receptor) include Primary Localized Cutaneous Amyloidosis and Primary Cutaneous Amyloidosis. This protein-coding gene is involved in key signaling pathways such as Akt signaling and interleukin-6 family signaling. Functionally, OSMR is linked to growth factor binding and oncostatin-M receptor activity, contributing to cellular communication and inflammatory responses. OSMR encodes a receptor that forms a complex with IL31RA to bind Oncostatin M (OSM), activating STAT3 and related pathways. Through these interactions, OSMR plays a key role in inflammatory signaling and hepatocellular carcinoma (HCC) progression by influencing liver inflammation, cancer stem cell differentiation, epithelial–mesenchymal transition, and tumor angiogenesis. Elevated OSMR signaling is associated with poor prognosis in HCC, especially in NAFLD/NASH-related cases [64, 65, 66].	PMID: 35064579
RSPO4	R-Spondin 4	Diseases linked to RSPO4 include Nonsyndromic Congenital Nail Disorder 4 and Tetraamelia Syndrome 2. This gene is also involved in key pathways such as WNT signaling and non-coding RNA-mediated Wnt signaling. RSPO4 functions as an activator of the canonical Wnt signaling pathway by binding to LGR4–6 receptors, which in turn interact with LRP6 and Frizzled receptors to promote Wnt signal transduction and target gene expression. Additionally, RSPO4 enhances both canonical (β-catenin-dependent) and non-canonical Wnt signaling by inhibiting ZNRF3, a negative regulator of this pathway.	Novel
CCL3L3	C-C Motif Chemokine Ligand 3 Like 3	Diseases associated with CCL3L3 include Autism Spectrum Disorder. Among its related pathways are Class A/1 (Rhodopsin-like receptors) and Cytokine Signaling in Immune system. CCL3 is a chemokine that attracts lymphocytes and monocytes by binding to CCR1, CCR3, and CCR5 receptors. It contributes to HCC progression by promoting angiogenesis and tumor-associated inflammation, primarily through the CCR1-CCL3 axis, which enhances MMP9 expression in immune cells [67].	PMID: 16284949

Continued on next page

Table S2 Continued from previous page

Gene Symbol	Gene Name	Known Functions	Evidence
TSHZ2	Teashirt Zinc Finger Homeobox 2	TSHZ2 is a transcription factor involved in developmental processes and acts as a tumor suppressor, particularly in breast tissue, with implications in other diseases such as interstitial lung disease.	Novel
CYP1A2	Cytochrome P450 Family 1 Subfamily A Member 2	Diseases associated with CYP1A2 include Major Depressive Disorder and Porphyria Cutanea Tarda. Among its related pathways are Clomipramine Pathway, Pharmacokinetics and Metapathway biotransformation Phase I and II. CYP1A2 is a cytochrome P450 enzyme primarily involved in the hepatic metabolism of endogenous hormones and xenobiotics. In the context of HCC, it plays a dual role: Carcinogen Metabolism, and Tumor Suppression [68].	PMID: 33500715
GATD3B	Glutamine Amidotransferase Class 1 Domain Containing 3	Diseases associated with GATD3 include Down Syndrome and Granulomatous Orchitis. Among its related pathways is Gastric cancer network 1.	Novel
GIPC2	GIPC PDZ Domain Containing Family Member 2	Diseases associated with GIPC2 include Waardenburg Syndrome, Type 4A and Pheochromocytoma. In certain cancers, such as pheochromocytoma/paraganglioma (PPGL) and colorectal cancer (CRC), GIPC2 functions as a tumor suppressor.	Novel
STMND1	Stathmin Domain Containing 1	Diseases associated with STMND1 include Myoclonic Epilepsy Of Lafora 1 and Spinocerebellar Ataxia 1. STMN1 is an oncogene highly expressed in HCC, promoting tumor proliferation, migration, drug resistance, and cancer stemness. It facilitates crosstalk with hepatic stellate cells via the HGF/MET signaling pathway, contributing to tumor progression and microenvironment remodeling [69].	PMID: 31785057
CNR1	Cannabinoid Receptor 1	Diseases associated with CNR1 include Cannabis Dependence and Cannabis Abuse. Among its related pathways are GPCR downstream signalling and Class A/1 (Rhodopsin-like receptors). CNR1 promotes hepatocellular carcinoma (HCC) progression by enhancing hepatocyte proliferation via FOXM1 induction and supporting an immunosuppressive tumor microenvironment through IDO2-mediated Treg activation. Its inhibition reduces tumor burden, highlighting its potential as a therapeutic target in HCC [70].	PMID: 25580584

Continued on next page

Table S2 Continued from previous page

Gene Symbol	Gene Name	Known Functions	Evidence
HOXA3	Homeobox A3	HOXA3 is involved in pathways such as RNA Polymerase I promoter activation and the development of the nervous system. HOX family members are broadly involved in regulating cell proliferation, the cell cycle, immune responses, and the tumor microenvironment in HCC. Aberrant expression and methylation of HOX genes, including HOXA3, may also hold diagnostic value in liver cancer [71].	PMID: 35251173

Table S3. The top 20 predicted results of HCC-related lncRNAs based on MTPrior with the summaries of their functions.

LncRNA Symbol	LncRNA Name	Known Functions	Evidence
HNF1A-AS1	HNF1A Antisense RNA 1	Diseases associated with HNF1A-AS1 include Esophagus Adenocarcinoma and Adenocarcinoma. HNF1A-AS1 promotes HCC progression by enhancing cell proliferation, cell cycle progression, and autophagy. It acts through epigenetic silencing of tumor suppressors (e.g., NKD1, p21) via EZH2 interaction and functions as a ceRNA by sponging miR-30b-5p, thereby upregulating targets like Bcl-2 and ATG5 [72, 73].	PMID: 28292020
PVT1	Pvt1 Oncogene	PVT1 is a cancer-associated lncRNA linked to breast, ovarian, and blood cancers. It promotes tumorigenesis by regulating MYC and is also associated with renal disease in diabetes and cleft lip. PVT1 is an oncogenic lncRNA upregulated in HCC, promoting tumor cell proliferation, invasion, and migration by regulating the miR-3619-5p/MKL1 and EZH2/miR-214 pathways, making it a potential therapeutic target [74].	Lnc2Cancer v3.0; PMID: 32156248
TTN-AS1	TTN Antisense RNA 1	Diseases associated with TTN-AS1 include Tibial Muscular Dystrophy and Third-Degree Atrioventricular Block. TTN-AS1 promotes HCC progression and sorafenib resistance by sponging miR-16-5p and miR-134, leading to upregulation of cyclin E1 and ITGB1, and activation of the PTEN/Akt pathway [75, 76].	PMID: 33387127; 33378944

Continued on next page

Table S3 Continued from previous page

LncRNA Symbol	LncRNA Name	Known Functions	Evidence
LINC00023	Maternally Expressed 3	Diseases associated with LINC00023 include Kagami-Ogata Syndrome and Liver Disease. The deregulation of linc00023 was reported to be involved in the progression of various tumours, including meningioma, hepatocellular cancer, and breast cancer. LINC00023 implicated in bladder cancer progression, though its specific function remains underexplored. It may regulate gene expression and influence tumor development, potentially through interactions with small nucleolar RNAs (snoRNAs) or as part of small nucleolar RNA host genes (SNHGs) [77].	PMID: 36831352
LINC01762	Long Intergenic Non-Protein Coding RNA 1762	LINC01762 were reported to be involved in the prognosis of colon adenocarcinoma, hepatocellular carcinoma and renal cell carcinoma. In HCC, the expression level of LINC01762 is significantly reduced in cancer cell lines, suggesting a potential tumor-suppressive role [78].	PMID: 36836069
LINC00294	Long Intergenic Non-Protein Coding RNA 294	Diseases associated with LINC00294 include Glioma Susceptibility 1 and HCC. LINC00294 is upregulated in HCC and promotes tumor progression by enhancing cell proliferation and aerobic glycolysis (Warburg effect), partly through m6A-mediated stabilization and regulation of HK2 and GLUT1 mRNA expression [79].	PMID: 38125436
LINC01620	Long Intergenic Non-Protein Coding RNA 1620	Diseases associated with LINC01620 include Myasthenic Syndrome, Congenital, 14. LINC01620 is a validated human long non-coding RNA located on chromosome 20q13.12.	RNADisease V4.0
LINC00265	Long Intergenic Non-Protein Coding RNA 265	Diseases associated with LINC00265 include Glioma Susceptibility 1. LINC00265 promotes HCC progression by binding to E2F1 and enhancing its recruitment to the CDK2 promoter, thereby upregulating CDK2 expression and driving cell proliferation, migration, and invasion [80].	PMID: 35892231
TMEM51-AS1	TMEM51 Antisense RNA 1	Diseases associated with TMEM51-AS1 include Chromophobe Renal Cell Carcinoma and Hepatocellular Carcinoma. Bioinformatic analysis indicated that TMEM51-AS1 expression is linked to immune evasion, RNA methylation, and DNA damage repair processes [81].	RNADisease V4.0

Continued on next page

Table S3 Continued from previous page

LncRNA Symbol	LncRNA Name	Known Functions	Evidence
HEPFAL	Hepatocellular Carcinoma Ferroptosis Associative LncRNA	HEPFAL is a ferroptosis-related lncRNA that is downregulated in HCC and promotes ferroptosis by reducing SLC7A11 expression, increasing lipid ROS and iron levels, and enhancing sensitivity to erastin, partly through mTORC1-related pathways [82].	PMID: 36008384
LINC01587	Long Intergenic Non-Protein Coding RNA 1587	Diseases associated with LINC01587 include Neuroblastoma and Ellis-Van Creveld Syndrome. LINC01587 is a long intergenic non-coding RNA that plays a role in regulating gene expression and other cellular functions.	Novel
TTTY14	Testis Expressed Transcript, Y-Linked 14	Diseases associated with TTTY14 include Invasive Bladder Transitional Cell Carcinoma. TTTY14 is a Y-linked lncRNA that acts as an oncogene in testicular germ cell tumors (TGCT), promoting tumor proliferation and serving as a poor prognostic marker. Its expression is linked to immune cell dysfunction and may impact drug sensitivity by modulating the tumor immune microenvironment [83].	Novel
Lnc-RNASEH2B-4 (DLEU1)	Deleted in Lymphocytic Leukemia 1	Diseases associated with DLEU1 include Ovary Epithelial Cancer and Leukemia, Chronic Lymphocytic. DLEU1 promotes HCC progression by sponging miR-133a, leading to upregulation of IGF-1R and activation of the PI3K/AKT pathway, thereby enhancing proliferation, migration, invasion, and EMT. Targeting the DLEU1/miR-133a/IGF-1R axis may offer therapeutic potential in HCC [84].	PMID: 31207081
Lnc-ZSCAN1-3 (ZNF606-AS1)	ZNF606 Antisense RNA 1	Diseases associated with ZNF606-AS1 include Adult Primary Hepatocellular Carcinoma, Adult Hcc, Adult Primary Carcinoma of The Liver Cell, and Adult Primary Hepatoma. It is a potential target for understanding and potentially treating hepatocellular carcinoma.	Novel
LINC01106	Long Intergenic Non-Protein Coding RNA 1106	Long non-coding RNA LINC01106 influences colorectal cancer cell proliferation and apoptosis via the STAT3 pathway. Silencing LINC01106 markedly reduced proliferation and increased apoptosis in SW480 cells [85].	Novel

Continued on next page

Table S3 Continued from previous page

LncRNA Symbol	LncRNA Name	Known Functions	Evidence
LINC00271 (AHI1-DT)	AHI1 Divergent Transcript	Diseases associated with AHI1-DT include Thyroid Cancer, Nonmedullary, 1 and Schizophrenia. LINC00271 acts as a potential tumor suppressor and independent risk factor associated with extrathyroidal extension, lymph node metastasis, advanced tumor stage, and recurrence. Its expression is significantly reduced in tumors compared to normal tissues, and low LINC00271 levels are linked to pathways involved in cell adhesion, cell cycle, p53, and JAK/STAT signaling, indicating its role in tumor progression and prognosis [86].	Novel
LINC00652	Long Intergenic Non-Protein Coding RNA 652	Diseases associated with LINC00652 include Glioma Susceptibility 1. LINC00652 suppresses the activation of the cAMP/PKA pathway by targeting GLP-1R, thereby diminishing the protective effects of sevoflurane against myocardial ischemia/reperfusion injury in mice [87].	Novel
ACTA2-AS1	ACTA2 Antisense RNA 1	ACTA2-AS1:4 acts as a suppressor in liver cancer, as its knockdown promotes proliferation, migration, and invasion of liver cancer cells [88].	Novel
LINC01559	Long Intergenic Non-Protein Coding RNA 1559	LINC01559 promotes HCC progression and oxaliplatin (L-OHP) resistance by acting as a competing endogenous RNA. It sponges miR-511 and miR-6783-3p, enhancing proliferation, migration, and drug resistance in HCC cells. High LINC01559 expression is associated with poor prognosis and may serve as a therapeutic target and biomarker for treatment response in HCC [89, 90].	PMID: 33588099; 32698745
LINC02544	Long Intergenic Non-Protein Coding RNA 2544	LINC02544 is a long non-coding RNA involved in regulating gene expression and participating in multiple cellular functions. It has been linked to the development and progression of several cancers, including breast and lung cancer, where it affects tumor growth, cell survival, and metastasis.	Novel

Supplementary References

[1] Li Y, Guo Z, Wang K, Gao X, Wang G. End-to-end interpretable disease-gene association prediction. Briefings in bioinformatics. 2023; 24(3).

[2] Tran VD, Sperduti A, Backofen R, Costa F. Heterogeneous networks integration for disease-gene prioritization with node kernels. Bioinformatics. 2020; 36(9):2649-2656.

[3] Xu L, Liang G, Liao C, Chen GD, Chang CC. k-Skip-n-Gram-RF: A Random Forest Based Method for Alzheimer's Disease Protein Identification. Frontiers in genetics. 2019; 10:33.

[4] Chen X, Yan GY. Novel human lncRNA-disease association inference based on lncRNA expression profiles. Bioinformatics. 2013;29(20):2617-24.

[5] Zhu R, Wang Y, Liu JX, Dai LY. IPCARF: improving lncRNA-disease association prediction using incremental principal component analysis feature selection and a random forest classifier. BMC Bioinformatics. 2021;22(1):175.

[6] Chen Q, Lai D, Lan W, Wu X, Chen B, Liu J, Chen YP, Wang J. ILDMSF: Inferring Associations Between Long Non-Coding RNA and Disease Based on Multi-Similarity Fusion. IEEE/ACM transactions on computational biology and bioinformatics. 2021; 18(3):1106-1112.

[7] Fu G, Wang J, Domeniconi C, Yu G. Matrix factorization-based data fusion for the prediction of lncRNA-disease associations. Bioinformatics. 2018; 34(9):1529-1537.

[8] Wang Y, Yu G, Wang J, Fu G, Guo M, Domeniconi C. Weighted matrix factorization on multi-relational data for LncRNA-disease association prediction. Methods. 2020;173:32-43.

[9] Wang Y, Yu G, Domeniconi C, Wang J, Zhang X, Guo M. Selective matrix factorization for multi-relational data fusion. Proceedings of the International Conference on Database Systems for Advanced Applications; 2019. p. 313–329.

[10] Lu C, Yang M, Luo F, Wu FX, Li M, Pan Y, Li Y, Wang J. Prediction of lncRNA-disease associations based on inductive matrix completion. Bioinformatics. 2018; 34(19):3357-3364.

[11] Deng L, Li W, Zhang J. LDAH2V: Exploring Meta-Paths Across Multiple Networks for lncRNA-Disease Association Prediction. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 2021; 18(4):1572-1581.

[12] Li Y, Li J, Bian N. DNILMF-LDA: Prediction of lncRNA-Disease Associations by Dual-Network Integrated Logistic Matrix Factorization and Bayesian Optimization. Genes. 2019; 10(8):608.

[13] Erten S, Bebek G, Koyutürk M. Vavien: an algorithm for prioritizing candidate disease genes based on topological similarity of proteins in interaction networks. Journal of computational biology. 2011; 18(11):1561-1574.

[14] Lysenko A, Boroevich KA, Tsunoda T. Arete - candidate gene prioritization using biological network topology with additional evidence types. BioData mining. 2017; 10:1-12.

[15] Sheng N, Wang Y, Huang L, Gao L, Cao Y, Xie X, Fu Y. Multi-task prediction-based graph contrastive learning for inferring the relationship among lncRNAs, miRNAs and diseases. Briefings in bioinformatics. 2023; 24(5).

[16] Peng J, Bai K, Shang X, Wang G, Xue H, Jin S, Cheng L, Wang Y, Chen J. Predicting disease-related genes using integrated biomedical networks. BMC Genomics. 2017; 18:1043.

[17] Xiao X, Zhu W, Liao B, Xu J, Gu C, Ji B, Yao Y, Peng L, Yang J. BPLLDA: Predicting lncRNA-Disease Associations Based on Simple Paths With Limited Lengths in a Heterogeneous Network. Frontiers in genetics. 2018;9:411.

[18] Sun J, Shi H, Wang Z, Zhang C, Liu L, Wang L, He W, Hao D, Liu S, Zhou M. Inferring novel lncRNA-disease associations based on a random walk model of a lncRNA functional similarity network. Molecular BioSystems. 2014; 10(8):2074-81.

[19] Gu C, Liao B, Li X, Cai L, Li Z, Li K, Yang J. Global network random walk for predicting potential human lncRNA-disease associations. Scientific reports. 2017; 7(1):12442.

[20] Zhao X, Yang Y, Yin M. MHRWR: Prediction of lncRNA-Disease Associations Based on Multiple Heterogeneous Networks. IEEE/ACM transactions on computational biology and bioinformatics. 2021; 18(6):2577-2585.

[21] Chen X, You ZH, Yan GY, Gong DW. IRWRLDA: improved random walk with restart for lncRNA-disease association prediction. Oncotarget. 2016; 7(36):57919-57931.

[22] Yu G, Fu G, Lu C, Ren Y, Wang J. BRWLDA: bi-random walks for predicting lncRNA-disease associations. Oncotarget. 2017;8(36):60429-60446.

[23] Wang L, Shang M, Dai Q, He PA. Prediction of lncRNA-disease association based on a Laplace normalized random walk with restart algorithm on heterogeneous networks. BMC Bioinformatics. 2022;23(1):5.

[24] Li Y, Kuwahara H, Yang P, Song L, Gao X. PGCN: Disease gene prioritization by disease and gene embedding through graph convolutional neural networks. biorxiv. 2019; 532226. doi: https://doi.org/10.1101/532226.

[25] Zhu L, Hong Z, Zheng H. Predicting gene-disease associations via graph embedding and graph convolutional networks. IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2019; pp. 382-389.

[26] Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907. 2016. Available from: https://doi.org/10.48550/arXiv.1609.02907

[27] Wang S, Hui C, Zhang T, Wu P, Nakaguchi T, Xuan P. Graph Reasoning Method Based on Affinity Identification and Representation Decoupling for Predicting lncRNA-Disease Associations. Journal of Chemical Information and Modeling. 2023; 63(21):6947-6958.

[28] Zhang W, Wei H, Zhang W, Wu H, Liu B. Multiple types of disease-associated RNAs identification for disease prognosis and therapy using heterogeneous graph learning. Science China Information Sciences. 2024; 67(8):1-2.

[29] Lan W, Wu X, Chen Q, Peng W, Wang J, Chen YP. GANLDA: Graph attention network for lncRNA-disease associations prediction. Neurocomputing. 2022; 469:384–393.

[30] Xuan P, Sheng N, Zhang T, Liu Y, Guo Y. CNNDLP. CNNDLP: A method based on convolutional autoencoder and convolutional neural network with adjacent edge attention for predicting lncRNA–disease associations. International Journal of Molecular Sciences, 2019; 20(17):4260.

[31] Zhang Z, Xu J, Wu Y, Liu N, Wang Y, Liang Y. CapsNet-LDA: predicting lncRNA-disease associations using attention mechanism and capsule network based on multi-view data. Briefings in Bioinformatics. 2023; 24(1).

[32] Lagisetty Y, Bourquard T, Al-Ramahi I, Mangleburg CG, Mota S, Soleimani S, Shulman JM, Botas J, Lee K, Lichtarge O. Identification of risk genes for Alzheimer's disease by gene embedding. Cell Genom. 2022; 2(9):100162.

[33] Perozzi B, Al-Rfou R, Skiena S. Deepwalk: Online learning of social representations. InProceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining 2014; pp. 701-710.

[34] Grover A, Leskovec J. node2vec: Scalable Feature Learning for Networks. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016; 2016:855-864.

[35] Velickovic P, Cucurull G, Casanova A, Romero A, Lio P, Bengio Y. Graph attention networks. stat. 2017; 1050(20):10-48550.

[36] Nelson W, Zitnik M, Wang B, Leskovec J, Goldenberg A, Sharan R. To Embed or Not: Network Embedding as a Paradigm in Computational Biology. Front Genet. 2019 ; 10:381.

[37] Wang X, Bo D, Shi C, Fan S, Ye Y, Philip SY. A survey on heterogeneous graph embedding: methods, techniques, applications and sources. IEEE Transactions on Big Data. 2022; 9(2):415-36.

[38] Chen L, Xuan J, Riggins RB, Clarke R, Wang Y. Identifying cancer biomarkers by network-constrained support vector machines. BMC systems biology. 2011; 5:1-20.

[39] Himmelstein DS, Baranzini SE. Heterogeneous Network Edge Prediction: A Data Integration Approach to Prioritize Disease-Associated Genes. PLoS computational biology. 2015; 11(7).

[40] Ghanat Bari M, Ung CY, Zhang C, Zhu S, Li H. Machine learning-assisted network inference approach to identify a new class of genes that coordinate the functionality of cancer networks. Scientific Reports, 2017; 7(1):6993.

[41] Zhou M, Wang X, Li J, Hao D, Wang Z, Shi H, Han L, Zhou H, Sun J. Prioritizing candidate disease-related long non-coding RNAs by walking on the heterogeneous lncRNA and disease network. Molecular BioSystems. 2015; 11(3):760-9.

[42] Xu C, Qi R, Ping Y, Li J, Zhao H, Wang L, Du MY, Xiao Y, Li X. Systemically identifying and prioritizing risk lncRNAs through integration of pan-cancer phenotype associations. Oncotarget. 2017; 8(7):12041-12051.

[43] Ding L, Wang M, Sun D, Li A. TPGLDA: Novel prediction of associations between lncRNAs and diseases via lncRNA-disease-gene tripartite graph. Sci Rep. 2018 Jan 18;8(1):1065. doi: 10.1038/s41598-018-19357-3.

[44] Zhang B, Wang H, Ma C, Huang H, Fang Z, Qu J. LDAGM: prediction lncRNA-disease asociations by graph convolutional auto-encoder and multilayer perceptron based on multi-view heterogeneous networks. BMC Bioinformatics. 2024 Oct 15;25(1):332.

[45] Yao D, Deng Y, Zhan X, Zhan X. Predicting lncRNA-disease associations using multiple metapaths in hierarchical graph attention networks. BMC Bioinformatics. 2024 Jan 29;25(1):46.

[46] Luo JP, Wang J, Huang JH. CDKN2A is a prognostic biomarker and correlated with immune infiltrates in hepatocellular carcinoma. *Biosci Rep.* 2021 Oct 29;41(10):BSR20211103. doi:10.1042/BSR20211103.

[47] Yang G, Wang G, Xiong Y, Sun J, Li W, Tang T, Li J. CDC20 promotes the progression of hepatocellular carcinoma by regulating epithelial-mesenchymal transition. Molecular medicine reports. 2021;24(1):483.

[48] Zhao S, Zhang Y, Lu X, Ding H, Han B, Song X, Miao H, Cui X, Wei S, Liu W, et al. CDC20 regulates the cell proliferation and radiosensitivity of P53 mutant HCC cells through the Bcl-2/Bax pathway. International journal of biological sciences. 2021;17(13):3608.

[49] Liao X, Bu Y, Jiang S, Chang F, Jia F, Xiao X, Song G, Zhang M, Ning P, Jia Q. CCN2–MAPK–Id-1 loop feedback amplification is involved in maintaining stemness in oxaliplatin-resistant hepatocellular carcinoma. Hepatology international. 2019;13:440–453.

[50] Pi L, Sun C, Jn-Simon N, Basha S, Thomas H, Figueroa V, Zarrinpar A, Cao Q, Petersen B. CCN2/CTGF promotes liver fibrosis through crosstalk with the Slit2/Robo signaling. Journal of Cell Communication and Signaling. 2023;17(1):137–150.

[51] Mai Y, Ji Z, Tan Y, Feng L, Qin J. BIRC5 knockdown ameliorates hepatocellular carcinoma progression via regulating PPARγ pathway and cuproptosis. Discover Oncology. 2024;15(1):706.

[52] Wang L, Li B, Bo X, Yi X, Xiao X, Zheng Q. Hypoxia-induced LncRNA DACT3-AS1 upregulates PKM2 to promote metastasis in hepatocellular carcinoma through the HDAC2/FOXA3 pathway. Experimental & molecular medicine. 2022;54(6):848–860.

[53] Li G, Zhu L, Guo M, Wang D, Meng M, Zhong Y, Zhang Z, Lin Y, Liu C, Wang J, et al. Characterisation of forkhead box protein A3 as a key transcription factor for hepatocyte regeneration. JHEP Reports. 2023;5(12):100906.

[54] Tan XL, Wang Z, Liao S, Yi M, Tao D, Zhang X, Leng X, Shi J, Xie S, Yang Y, et al. NR0B1 augments sorafenib resistance in hepatocellular carcinoma through promoting autophagy and inhibiting apoptosis. Cancer Science. 2024;115(2):465–476.

[55] Zhang X, Xiao Z, Zhang X, Li N, Sun T, Zhang J, Kang C, Fan S, Dai L, Liu X. Signature construction and molecular subtype identification based on liver-specific genes for prediction of prognosis, immune activity, and anti-cancer drug sensitivity in hepatocellular carcinoma. Cancer Cell International. 2024;24(1):78.

[56] Xiang S, Li J, Shen J, Zhao Y, Wu X, Li M, Yang X, Kaboli PJ, Du F, Zheng Y, et al. Identification of prognostic genes in the tumor microenvironment of hepatocellular carcinoma. Frontiers in Immunology. 2021;12;653836.

[57] Liang X, Liu Q, Zhu S, Li Z, Chen H, Su Z. GSDME has prognostic and immunotherapeutic significance in residual hepatocellular carcinoma after insufficient radiofrequency ablation. Translational Oncology. 2024;39:101796.

[58] Chen S, Zhang P, Zhu G, Wang B, Cai J, Song L, Wan J, Yang Y, Du J, Cai Y, et al. Targeting GSDME-mediated macrophage polarization for enhanced antitumor immunity in hepatocellular carcinoma. Cellular & Molecular Immunology. 2024:1–17.

[59] Lu Y, Xu J, Lin H, Zhu M, Li M. Gasdermin E mediates pyroptosis in the progression of hepatocellular carcinoma: a double-edged sword. Gastroenterology Report. 2024;12:goae102.

[60] Liu K, Zhang Y, Zhang C, Zhang Q, Li J, Xiao F, Li Y, Zhang R, Dou D, Liang J, et al. Methylation of S100A8 is a promising diagnosis and prognostic marker in hepatocellular carcinoma. Oncotarget. 2016;7(35):56798.

[61] De Ponti A, Wiechert L, Schneller D, Pusterla T, Longerich T, Hogg N, Vogel A, Schirmacher P, Hess J, Angel P. A pro-tumorigenic function of S100A8/A9 in carcinogen-induced hepatocellular carcinoma. Cancer Lett. 2015;369(2):396–404.

[62] Vilarinho S, Erson-Omay EZ, Harmanci AS, Morotti R, Carrion-Grant G, Baranoski J, Knisely AS, Ekong U, Emre S, Yasuno K, et al. Paediatric hepatocellular carcinoma due to somatic CTNNB1 and NFE2L2 mutations in the setting of inherited bi-allelic ABCB11 mutations. J Hepatol. 2014;61(5):1178–1183.

[63] Chen H. P049: The expression profile of ABCB11 gene in hepatocellular carcinoma and its association with clinical outcomes. Genetics in Medicine Open. 2023;1(1):100068.

[64] Yang X, Shao C, Duan L, Hou X, Huang Y, Gao L, Zong C, Liu W, Jiang J, Ye F, et al. Oncostatin M promotes hepatic progenitor cell activation and hepatocarcinogenesis via macrophage-derived tumor necrosis factor-α. Cancer Lett. 2021;517:46–54.

[65] Yamashita T, Honda M, Nio K, Nakamoto Y, Yamashita T, Takamura H, Tani T, Zen Y, Kaneko S. Oncostatin m renders epithelial cell adhesion molecule–positive liver cancer stem cells sensitive to 5-fluorouracil by inducing hepatocytic differentiation. Cancer Res. 2010;70(11):4687–4697.

[66] Di Maira G, Foglia B, Napione L, Turato C, Maggiora M, Sutti S, Novo E, Alvaro M, Autelli R, Colombatto S, et al. Oncostatin M is overexpressed in NASH-related hepatocellular carcinoma and promotes cancer cell invasiveness and angiogenesis. The Journal of pathology. 2022;257(1):82–95.

[67] Yang X, Lu P, Fujii C, Nakamoto Y, Gao J-L, Kaneko S, Murphy PM, Mukaida N. Essential contribution of a chemokine, CCL3, and its receptor, CCR1, to hepatocellular carcinoma progression. Int J Cancer. 2006;118(8):1869–1876.

[68] Yu J, Xia X, Dong Y, Gong Z, Li G, Chen GG, San Lai PB. CYP1A2 suppresses hepatocellular carcinoma through antagonizing HGF/MET signaling. Theranostics. 2021;11(5):2123.

[69] Zhang R, Gao X, Zuo J, Hu B, Yang J, Zhao J, Chen J. STMN1 upregulation mediates hepatocellular carcinoma and hepatic stellate cell crosstalk to aggravate cancer by triggering the MET pathway. Cancer science. 2020;111(2):406-417.

[70] Mukhopadhyay B, Schuebel K, Mukhopadhyay P, Cinar R, Godlewski G, Xiong K, Mackie K, Lizak M, Yuan Q, Goldman D, et al. Cannabinoid receptor 1 promotes hepatocellular carcinoma initiation and progression through multiple mechanisms. Hepatology. 2015;61(5):1615-1626.

[71] Jin Z, Sun D, Song M, Zhu W, Liu H, Wang J, Shi G. Comprehensive analysis of HOX family members as novel diagnostic and prognostic markers for hepatocellular carcinoma. Journal of Oncology. 2022;2022(1):5758601.

[72] Wang C, Mou L, Chai H-X, Wang F, Yin Y-Z, Zhang X-Y. Long non-coding RNA HNF1A-AS1 promotes hepatocellular carcinoma cell proliferation by repressing NKD1 and P21 expression. Biomedicine and Pharmacotherapy. 2017;89:926-932.

[73] Liu Z, Wei X, Zhang A, Li C, Bai J, Dong J. Long non-coding RNA HNF1A-AS1 functioned as an oncogene and autophagy promoter in hepatocellular carcinoma through sponging hsa-miR-30b-5p. Biochemical and Biophysical Research Communications. 2016;473(4):1268-1275.

[74] Liu H, Yin Y, Liu T, Gao Y, Ye Q, Yan J, Ha F. Long non-coding RNA PVT1 regulates the migration of hepatocellular carcinoma HepG2 cells via miR-3619-5p/MKL1 axis. Bosnian Journal of Basic Medical Sciences. 2021;21(2):187.

[75] Zhou Y, Huang Y, Dai T, Hua Z, Xu J, Lin Y, Han L, Yue X, Ho L, Lu J, et al. LncRNA TTN-AS1 intensifies sorafenib resistance in hepatocellular carcinoma by sponging miR-16-5p and upregulation of cyclin E1. Biomedicine and Pharmacotherapy. 2021;133:111030.

[76] Huang Y, Chu P, Bao G. Silencing of long non-coding RNA TTN-AS1 inhibits hepatocellular carcinoma progression by the microRNA-134/ITGB1 axis. Digestive Diseases and Sciences. 2021;1–13.

[77] Wang H, Feng Y, Zheng X, Xu X. The diagnostic and therapeutic role of snoRNA and lincRNA in bladder cancer. Cancers. 2023;15(4):1007.

[78] Liang G-Z, Wen X-F, Song Y-W, Zhang Z-J, Chen J, Chen Y-L, Pan W-D, He X-W, Hu T, Xian Z-Y. Construction and validation of a novel prognosis model in colon cancer based on cuproptosis-related long non-coding RNAs. Journal of Clinical Medicine. 2023;12(4):1528.

[79] Zhang R, Yang R, Huang Z, Xu X, Lv S, Guan X, Li H, Wu J. METTL3/YTHDC1-mediated upregulation of LINC00294 promotes hepatocellular carcinoma progression. Heliyon. 2023;9(12).

[80] Ge B, Zhang X, Zhou W, Mo Y, Su Z, Xu G, Chen Q. Linc00265 promotes metastasis and progression of hepatocellular carcinoma by interacting with E2F1 at the promoter of CDK2. Cell J (Yakhteh). 2022;24(6):294.

[81] Wu D, Xia Q, Su X, Mao Y, Mao J, Ding Q, Liu J, Zhong W, Zhang X, Li H, et al. Long non-coding RNA TMEM51-AS1 inhibits colorectal cancer progression. Discover Oncology. 2025;16(1):1-15.

[82] Zhang B, Bao W, Zhang S, Chen B, Zhou X, Zhao J, Shi Z, Zhang T, Chen Z, Wang L, et al. LncRNA HEPFAL accelerates ferroptosis in hepatocellular carcinoma by regulating SLC7A11 ubiquitination. Cell Death and Disease. 2022;13(8):734.

[83] Cao J, Liu L, Xue L, Luo Y, Liu Z, Guo J. Long non-coding RNA TTTY14 promotes cell proliferation and functions as a prognostic biomarker in testicular germ cell tumor. Heliyon. 2023;9(5).

[84] Zhang W, Liu S, Liu K, Liu Y. Long non-coding RNA deleted in lymphocytic leukaemia 1 promotes hepatocellular carcinoma progression by sponging miR-133a to regulate IGF-1R expression. Journal of Cellular and Molecular Medicine. 2019;23(8):5154–5164.

[85] Gu Y, Huang Y, Sun Y, Liang X, Kong L, Liu Z, Wang L. Long non-coding RNA LINC01106 regulates colorectal cancer cell proliferation and apoptosis through the STAT3 pathway. Nan Fang Yi Ke Da Xue Xue Bao = Journal of Southern Medical University. 2020;40(9):1259–1264.

[86] Ma B, Liao T, Wen D, Dong C, Zhou L, Yang S, Wang Y, Ji Q. Long intergenic non-coding RNA 271 is predictive of a poorer prognosis of papillary thyroid cancer. Scientific Reports. 2016;6(1):36973.

[87] Zhang S-B. Suppression of long non-coding RNA LINC00652 restores sevoflurane-induced cardioprotection against myocardial ischemia-reperfusion injury by targeting GLP-1R through the cAMP/PKA pathway in mice. Cellular Physiology and Biochemistry. 2018;49(4):1476–1491.

[88] Zhou, R.-J. and Lv, H.-Z. (2019). Knockdown of ACTA2-AS1 promotes liver cancer cell proliferation, migration and invasion. *Molecular Medicine Reports*, 19(3), 2263–2270.

[89] Su, Q. and Wang, H. (2021). Long non-coding RNA 01559 mediates the malignant phenotypes of hepatocellular carcinoma cells through targeting miR-511. *Clinics and Research in Hepatology and Gastroenterology*, 45(2), 101648.

[90] Dong, S., Fu, Y., Yang, K., Zhang, X., Miao, R., Long, Y., and Liu, C. (2021). Linc01559 served as a potential oncogene and promoted resistance of hepatocellular carcinoma to oxaliplatin by directly sponging miR-6783-3p. *Anti-Cancer Agents in Medicinal Chemistry*, 21(2), 278–286.