**A multi-scale neighbor topology guided transformer and Kolmogorov-Arnold network enhanced feature learning model for disease-related circRNA prediction**

Ping Xuan, Haoyuan Li, Hui Cui, Zelong Xu, Toshiya Nakaguchi, and Tiangang Zhang∗

1. **Calculation process of disease similarity**

Disease denominations in this context are based on the MeSH database (http://www.ncbi.nlm.nih.gov/), which provides a systematic disease classification framework structured as a directed acyclic graph (DAG). In this framework, nodes represent diseases and edges represent hierarchical or associative relationships between diseases. Disease A can be represented within a DAG, denoted as = (A，，), where is the set of nodes including node A and its ancestors in the MeSH hierarchy, and is the set of directed edges representing relationships between these nodes. For a node , we define the semantic contribution of disease to disease A in as ,

,

where is the semantic contribution factor set by the edges , and connects disease and its child disease . The semantic value of disease A is defined as ,

and the semantic similarity between two diseases is defined as,

.

1. **Determination process of the number of maximum scales**

The maximum scale of the multi-scale neighbor topology is denoted as *r*. To evaluate its impact on circRNA-disease association prediction performance, we test values of *r* from {1, 2, 3, 4, 5}. As shown in Supplementary table 1, our model achieves the optimal performance in terms of the AUC and AUPR metrics when *r* = 2. However, the model's prediction performance declines when *r* = 3, 4, and 5. This may be due to noise from including more neighbors.

**Supplementary table 1.** Prediction performance comparison for varying maximum scales of the multi-scale neighbor topology.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *r* | 1 | 2 | 3 | 4 | 5 |
| AUC | 0.944 | **0.946** | 0.934 | 0.930 | 0.924 |
| AUPR | 0.248 | **0.267** | 0.261 | 0.250 | 0.234 |

1. **Determination process of the number of restart probability**

The restart probability of the random walk, denoted by λ, is varied in experiments using values from the set {0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9}. According to Supplementary table 2, the model achieves the highest AUC and AUPR values when λ = 0.7. Specifically, AUC shows a variation with a range of 0.029 and a standard deviation of 0.008, while the AUPR has a variation range of 0.046 and a standard deviation of 0.013. As the value of λ decreases, the walking range increases. When λ is smaller than 0.7, the model's prediction performance declines. This may occur because a larger walking range incorporates information from less relevant distant nodes, introducing noise that degrades performance. When λ exceeds 0.7, the walking range shrinks, focusing more on the starting node, which limits the information from distant nodes.

**Supplementary table 2.** Prediction performance comparison for varying restart probability of the random walk.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| λ | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
| AUC | 0.917 | 0.931 | 0.929 | 0.932 | 0.937 | 0.940 | **0.946** | 0.940 | 0.939 |
| AUPR | 0.221 | 0.235 | 0.232 | 0.240 | 0.241 | 0.254 | **0.267** | 0.250 | 0.242 |

1. **Determination process of the number of attention heads**

The number of attention heads is denoted as . To evaluate the impact of the number of attention heads on circRNA-disease association prediction performance, we conduct experiments with values selected from {1, 2, 4, 6, 8}. The experimental results (Supplementary Table 3) show that our model achieve optimal performance in terms of AUC and AUPR metrics when .

**Supplementary table 3.** Prediction performance comparison for varying numbers of attention heads.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 4 | 6 | 8 |
| AUC | 0.937 | 0.942 | **0.946** | 0.940 | 0.937 |
| AUPR | 0.226 | 0.256 | **0.267** | 0.246 | 0.255 |

1. **Determination process of KAN neurons**

To evaluate the impact of different numbers of neurons in the KAN on circRNA-disease association prediction performance, we select combinations from the first-layer neuron counts {512, 1024, 2048, 4096} and the second-layer neuron counts {64, 128, 256, 512}. The experimental results (Supplementary Table 4) show that when the first layer has 1024 neurons and the second layer has 256 neurons, our model achieves the highest performance in terms of AUC and AUPR metrics.

**Supplementary table 4.** Prediction performance comparison for KAN networks with different numbers of neurons.

|  |  |  |  |
| --- | --- | --- | --- |
| 1st layer | 2nd layer | AUC | AUPR |
| 1024 | 64 | 0.927 | 0.237 |
| 1024 | 128 | 0.931 | 0.257 |
| 1024 | 256 | **0.946** | **0.267** |
| 1024 | 512 | 0.932 | 0.246 |
| 512 | 256 | 0.921 | 0.239 |
| 512 | 128 | 0.925 | 0.242 |
| 512 | 64 | 0.919 | 0.232 |
| 2048 | 512 | 0.934 | 0.247 |
| 2048 | 256 | 0.938 | 0.250 |
| 4096 | 512 | 0.932 | 0.235 |
| 4096 | 256 | 0.929 | 0.229 |

1. **Analysis of the contribution of pairwise global and local feature**

To analyze the contribution of pairwise global and local features to the performance of circRNA-disease association prediction, we collect the final values of the learnable parameters and for paired global and local features after training under five-fold cross-validation. The experimental results show that the values of and are both close to 0.5, indicating that the contributions of paired global and local features are balanced (Supplementary Table 5) . Both are equally important and provide valuable information to the model.

**Supplementary table 5.** Values of parameters and obtained from training.

|  |  |  |
| --- | --- | --- |
| Fold |  |  |
| 1 | 0.501 | 0.499 |
| 2 | 0.500 | 0.500 |
| 3 | 0.500 | 0.500 |
| 4 | 0.498 | 0.502 |
| 5 | 0.500 | 0.500 |

1. **Analysis of the impact of learnable parameters and on model performance stability**

We fix all other learnable parameters’ initial values, randomly initialize and , and update these parameters during back propagation through the loss function. To assess the impact of their random initialization on model’s performance, we train and test the MKCD model with different initialization values for and , calculating ten sets of AUC and AUPR values. As shown in Supplementary Table 6, random initialization of and causes small performance variations, with AUC varying by up to 0.006 (standard deviation 0.0021), and AUPR by up to 0.008 (standard deviation 0.0027). The experimental results indicate that MKCD exhibits high stability with respect to the initial values of and .

**Supplementary table 6.** Model performance under different initialized learnable weight parameters of and .

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| AUC | 0.943 | 0.946 | 0.944 | 0.943 | 0.943 | 0.946 | 0.949 | 0.947 | 0.945 | 0.948 |
| AUPR | 0.263 | 0.268 | 0.267 | 0.263 | 0.264 | 0.267 | 0.270 | 0.262 | 0.264 | 0.269 |

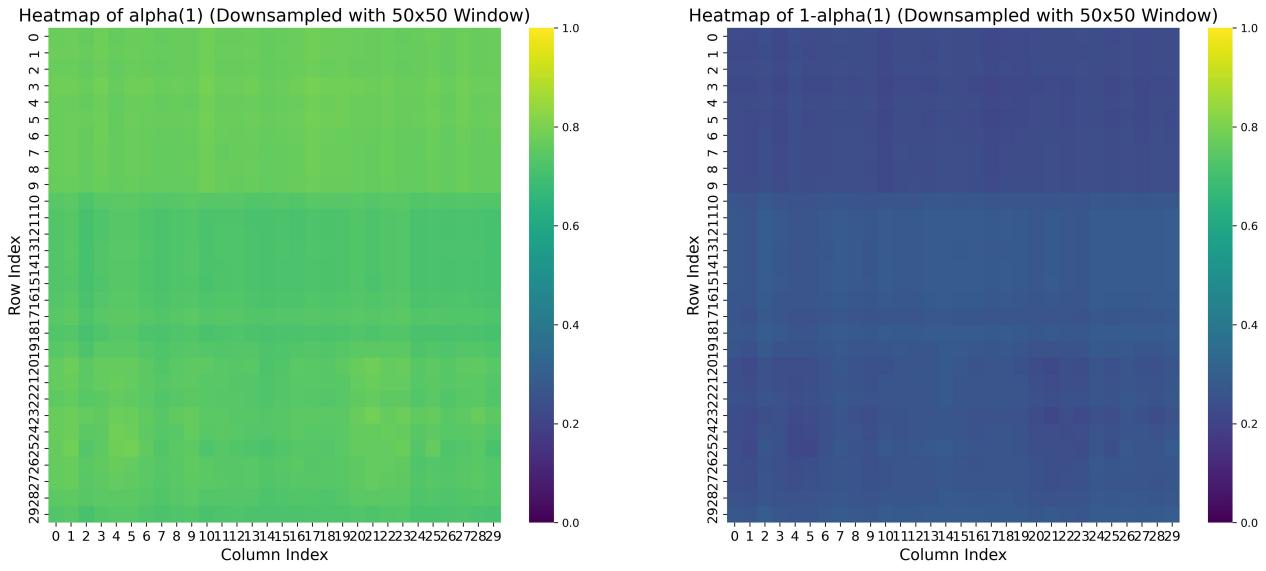
1. **Determination process of the number of DMTT Layers**

The number of DMTT (dynamic multi-scale neighbor topology-guided transformer) layers is denoted as . To evaluate the impact of the number of DMTT layers on circRNA-disease association prediction performance, we select values of from {1, 2, 3, 4, 5} for the experiment. The experimental results show that our model achieves the optimal performance in terms of both AUC and AUPR metrics when (supplementary Table 7).

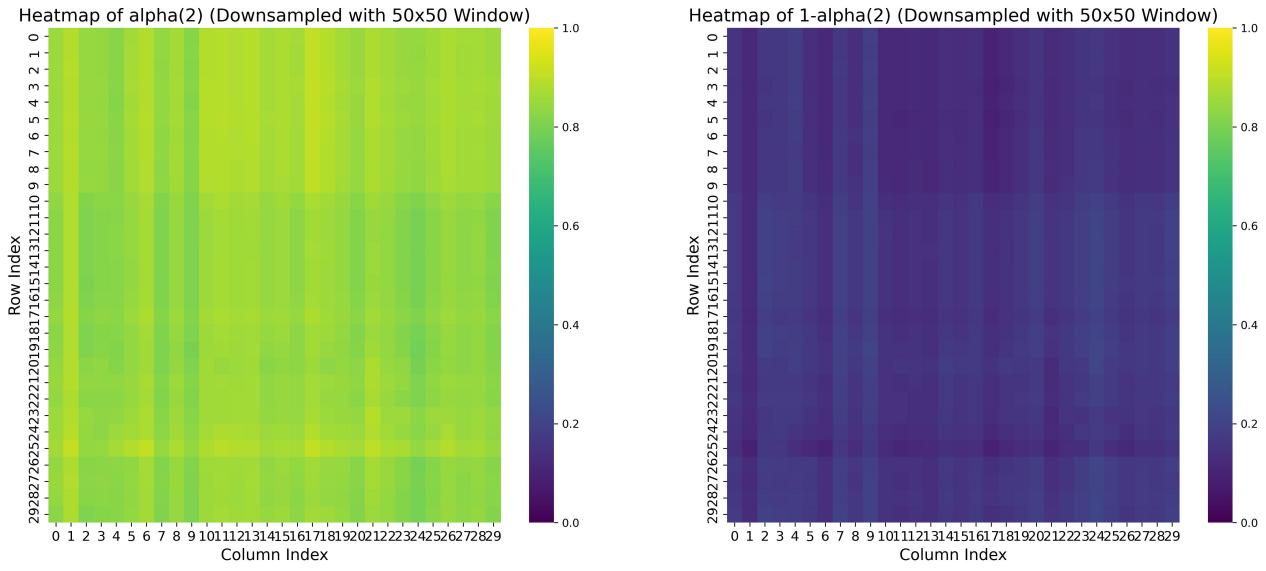
**Supplementary table 7.** Prediction performance comparison for varying layer numbers of DMTT.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 |
| AUC | 0.936 | **0.946** | 0.930 | 0.929 | 0.928 |
| AUPR | 0.231 | **0.267** | 0.264 | 0.252 | 0.248 |

To investigate the relative contributions of and in the fusion process of DMTT, we generate two sets of heatmaps based on the matrices , (1 - ), , and (1 - ) (Supplementary Fig 1 and 2). For each matrix, a 50×50 window is down-sampled, and the average value of the elements within the window is calculated as the pixel value for the heatmap. As shown in Supplementary Figure 1, after the first DMTT layer , the weight matrix corresponding to the encoded feature , exhibits values primarily ranging from 0.6 to 0.8, whereas the weight matrix for the original feature (i.e., 1 - ) ranges from 0.2 to 0.4. This indicates that the importance of is greater than that of . Similarly, as shown in Supplementary Fig 2, in the second layer of DMTT, the weight matrix corresponding to the encoded feature after two encodings, has values predominantly between 0.7 and 0.9, while the weight matrix for which has only been encoded once, ranges from 0.1 to 0.3. This indicates that the importance of is greater than that of . In summary, for a two-layer DMTT, plays a dominant role, while enhances the model's expressive power by providing supplementary features.



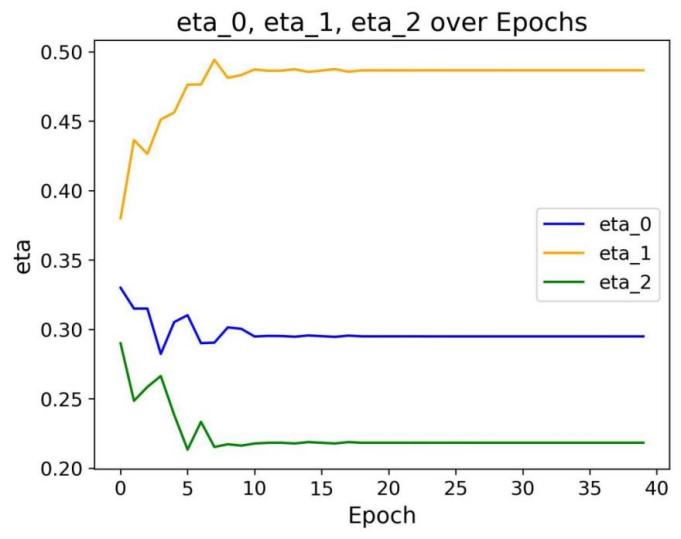
**Supplementary Fig 1.**



**Supplementary Fig 2.** and (1 - ).

1. **Training variations of learnable parameters , , and**

The learnable parameters for scales 0 to 2 are denoted as , , and , and they are randomly initialized. For each batch of data, we calculate the loss function for circRNA-disease association prediction and update the values of , , and using backpropagation. The training process consists of 40 epochs. As shown in Supplementary Figure 3, the values of , , and fluctuate significantly during the first 10 epochs, gradually converge between epochs 10 and 20, and stabilize after 20 epochs. The final values of , , and stabilize at 0.295, 0.487, and 0.218, respectively. The experimental results indicate that the weight proportion of the 1-scale neighbor topology embedding is the largest, which is likely because it includes neighbors that are directly connected to the target node, resulting in stronger associations. In contrast, the 2-scale neighbor topology has a smaller weight proportion, possibly due to its indirect, two-step connections to the target node.



**Supplementary figure 3.** Learnable parameters , and changes with epochs.

1. **Comparison of multi-scale and single-scale neighbor topologies**

To verify whether the multi-scale neighbor topology in MKCD outperforms single-order neighbor topologies in prediction performance, we test the model's prediction performance using only the *i*-th (*i* = 1, 2, 3) order neighbor topology. As shown in Supplementary Table 8, when using the multi-scale fused neighbor topology, our model achieves the optimal performance in both AUC and AUPR, with values of 0.946 and 0.267, respectively. Compared to our model, using only the 1-order neighbors results in a 0.4% and 1.7% decrease in AUC and AUPR, respectively; using only the 2-order neighbors leads to a 0.9% and 2.5% decrease, while using only the 3-order neighbors causes a 2.5% and 3.7% drop. The experimental results demonstrate that by adaptively weighted fusion of multi-order neighbor topologies, MKCD can effectively capture complementary information from different order neighbors, thereby improving the model's prediction performance.

**Supplementary table 8.** Prediction performance comparison for using different scale neighbor topologies.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Only-1 | Only-2 | Only-3 | Multi |
| AUC | 0.942 | 0.937 | 0.921 | **0.946** |
| AUPR | 0.250 | 0.242 | 0.230 | **0.267** |

1. **Comparison of different neighbor topology generation strategies**

To compare the performance differences between different neighbor topology generation strategies, we use Breadth-First Search (BFS) and Personalized PageRank as comparison methods, denoted as MKCD\_BFS and MKCD\_PR, respectively. The experimental results (Supplementary Table 9) show that the RWR-based AMNE module performs the best in terms of AUC and AUPR, with values of 0.946 and 0.267, respectively. Compared to MKCD\_BFS, our model improves AUC by 2.9% and AUPR by 5.0%; compared to MKCD\_PR, our model improves AUC by 1.0% and AUPR by 2.4%. The superior performance of RWR is attributed to its restart mechanism, which effectively balances local and global topology information, making it suitable for capturing the complex multi-scale relationships in the circRNA-disease association network.

**Supplementary table 9.** Prediction performance comparison for different neighbor topology generation strategies.

|  |  |  |  |
| --- | --- | --- | --- |
|  | MKCD | MKCD\_BFS | MKCD\_PR |
| AUC | **0.946** | 0.917 | 0.936 |
| AUPR | **0.267** | 0.217 | 0.243 |

1. **Determination process of the number and order of spline functions**

The number of spline functions is denoted as . To evaluate the impact of on circRNA-disease association prediction performance, we select values for from {3, 4, 5, 6, 7} for the experiment. The experimental results are shown in Supplementary Table 10. When = 5, our model achieves the optimal performance in both AUC and AUPR metrics. The variation in results in an AUC change range of 0.02, with a standard deviation of 0.0073, and an AUPR change range of 0.031, with a standard deviation of 0.0105.

The order of the spline functions affects the smoothness of the function. The order is denoted as *k*, and the range of *k* is {1, 2, 3, 4}. When *k* = 3, the model achieves the highest values of AUC and AUPR (Supplementary Table 11). The variation in *k* results in an AUC change range of 0.019, with a standard deviation of 0.0078, and an AUPR change range of 0.055, with a standard deviation of 0.0201. The experimental results indicate that the order of the spline functions, *k*, is more sensitive to the final performance than .

**Supplementary table 10.** Prediction performance comparison for different numbers of spline functions.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 3 | 4 | 5 | 6 | 7 |
| AUC | 0.926 | 0.940 | **0.946** | 0.946 | 0.941 |
| AUPR | 0.236 | 0.249 | **0.267** | 0.253 | 0.260 |

**Supplementary table 11.** Comparison of prediction performance with spline function of different orders.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 |
| AUC | 0.927 | 0.930 | **0.946** | 0.941 |
| AUPR | 0.212 | 0.235 | **0.267** | 0.250 |

1. **The impact of RWR and KAN on model training and inference time**

To evaluate whether RWR (random walk with restart) significantly increases the training and inference time of the model, we compare the model with and without RWR, where the model without RWR is denoted as MKCD\_noRWR. We also replace the KAN network in the pairwise global feature learning module with MLP, denoted as MKCD\_noKAN, to analyze the time cost of KAN. The training time is defined as the time for one epoch, and inference time is the time predict whether a pair of circRNA-disease is associated. As shown in Supplementary Table 12, incorporating RWR adds 0.061 seconds to training time per epoch (an 8.63% increase) and 0.016 ms to inference time per circRNA-disease pair. The full MKCD model with KAN and RWR increases training time for 40 epochs by approximately 0.105 seconds compared to MKCD\_noKAN (a 15.84% increase) and requires 0.028 ms more for inference per circRNA-disease pair. These results indicate that RWR and KAN only slightly increased the model's training and inference time.

**Supplementary table 12.** Comparison of training and inference time for MKCD and its variants.

|  |  |  |  |
| --- | --- | --- | --- |
|  | MKCD | MKCD\_noRWR | MKCD\_noKAN |
| Training time per epoch (s) | 0.768 | 0.707 | 0.663 |
| Inference time per circRNA-disease pair  (ms) | 0.352 | 0.336 | 0.324 |
| Training time increase (%) | - | 8.63 | 15.84 |

1. **Estimation of training/inference time for MKCD on large-scale datasets**

Pokec is a large-scale social network dataset containing 1,632,803 user nodes and 30,622,564 edges. We used an H100 GPU to train the model, with training time measured per epoch, and inference time measured per node pair prediction. As shown in Supplementary Table 13, our model takes approximately 1.85 hours to train one epoch on the Pokec dataset and 2.94 seconds for inference per node pair. The analysis indicates that MKCD demonstrates efficient training and reference performance on large-scale datasets.

**Supplementary table 13.** Comparison of training and inference time across datasets.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Node | Edge | Training time per epoch  (s) | | Inference time per node pair  (s) |
| CircRNA-disease | 1,527 | 2,728 | | 0.02 | 1.07e-5 |
| Pokec | 1,632,803 | 30,622,564 | | 6667.52 | 2.94 |

**Supplementary table 14.** Time complexity and training and inference times of MKCD and six comparison methods.

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Time complexity | Training time (s) | Inference time |
| SGFCCDA | O() | 39.42 | 0.462 |
| MLNGCF | O() | 32.14 | 0.423 |
| MDGF-MCEC | O() | 13.69 | 0.291 |
| Bi-SGTAR | O(m·+n·) | 8.28 | 0.068 |
| GraphCDA | O() | 6.67 | 0.177 |
| MPCLCDA | O() | 47.90 | 0.568 |
| MKCD | O() | 30.72 | 0.352 |

**Supplementary table 15.** Ablation experiment results for MKCD model on hepatocellular carcinoma.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| AMNE | DMTT | FGN | ACK | AUC | AUPR |
| **×** | **√** | **√** | **√** | 0.922 | 0.134 |
| **√** | **×** | **×** | **√** | 0.908 | 0.117 |
| **√** | **√** | **×** | **√** | 0.927 | 0.151 |
| **√** | **√** | **√** | **×** | 0.934 | 0.155 |
| **√** | **√** | **√** | **√** | **0.951** | **0.172** |

**Supplementary table 16.** Ablation experiment results for MKCD model on gastric cancer.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| AMNE | DMTT | FGN | ACK | AUC | AUPR |
| **×** | **√** | **√** | **√** | 0.928 | 0.128 |
| **√** | **×** | **×** | **√** | 0.911 | 0.105 |
| **√** | **√** | **×** | **√** | 0.932 | 0.136 |
| **√** | **√** | **√** | **×** | 0.941 | 0.150 |
| **√** | **√** | **√** | **√** | **0.953** | **0.169** |

**Supplementary table 17.** Ablation experiment results for MKCD model on glioma.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| AMNE | DMTT | FGN | ACK | AUC | AUPR |
| **×** | **√** | **√** | **√** | 0.919 | 0.092 |
| **√** | **×** | **×** | **√** | 0.912 | 0.089 |
| **√** | **√** | **×** | **√** | 0.937 | 0.106 |
| **√** | **√** | **√** | **×** | 0.940 | 0.121 |
| **√** | **√** | **√** | **√** | **0.947** | **0.135** |

**Supplementary table 18.** Results of the paired Wilcoxon test comparing MKCD with all other methods.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *p*-value | SGFCCDA | MLNGCF | MDGF-MCEC | Bi-SGTAR | GraphCDA | MPCLCDA |
| AUC | 2.71e-111 | 7.20e-58 | 2.13e-109 | 3.27e-60 | 6.06e-108 | 1.22e-76 |
| AUPR | 6.02e-109 | 6.60e-115 | 1.18e-41 | 1.96e-09 | 2.86e-26 | 1.29e-113 |
| F1 score | 2.20e-56 | 4.38e-98 | 4.72e-30 | 2.57e-08 | 1.52e-04 | 3.43e-79 |
| Precision | 3.66e-8 | 4.22e-114 | 1.80e-47 | 2.47e-20 | 3.24e-33 | 2.52e-103 |

**Supplementary table 19.** Top 15 predicted results of glioma-related circRNAs based on MKCD.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Rank | CircRNA name | Evidence | Rank | CircRNA name | Evidence |
| 1 | circ\_002136 | , PMID:30736838  PMID:30341906  , PMID:31511040  , PMID:32196629  , PMID:31669648  , PMID:28219405  , PMID:31038801  , PMID:30010402 | 9 | hsa\_circ\_0079593 | , PMID:31148222 |
| 2 | hsa\_circ\_0061868 | 10 | circSMO742 | , PMID:31895689 |
| 3 | circPTN | 11 | hsa\_circ\_0012129 | , PMID:29686222 |
| 4 | circ-TTBK2 | 12 | hsa\_circ\_0088732 | , PMID:32154171 |
| 5 | circ-EZH2 | 13 | circNFIX | , PMID:30072869 |
| 6 | hsa\_circ\_0000594 | 14 | circ-PTPRZ1 | , PMID:31364003 |
| 7 | hsa\_circ\_0005198 | 15 | hsa\_circ\_0014359 | , PMID:30745107 |
| 8 | hsa\_circ\_0000177 |  |  |  |

: circRNADisease v2.0.

**Supplementary table 20.** Top 15 predicted results of systemic lupus erythematosus-related circRNAs based on MKCD.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Rank | CircRNA name | Evidence | Rank | CircRNA name | Evidence |
| 1 | hsa\_circ\_0046599 | , PMID:29360436  , PMID:29360436  , PMID:29360436  , PMID:29360436  , PMID:31608065  , PMID:30628013  , PMID:30871426  , PMID:29700819 | 9 | hsa\_circ\_0008615 | , PMID:29360436 |
| 2 | hsa\_circ\_0001866 | 10 | hsa\_circ\_0021549 | , PMID:29360436 |
| 3 | hsa\_circ\_0034398 | 11 | has\_circ\_0049220 | , PMID:29606700 |
| 4 | hsa\_circ\_0003146 | 12 | hsa\_circ\_0092374 | , PMID:29360436 |
| 5 | hsa\_circ\_0000479 | 13 | hsa\_circ\_0040705 | , PMID:29360436 |
| 6 | hsa\_circ\_0057762 | 14 | hsa\_circ\_0012919 | , PMID:30237316 |
| 7 | circPTPN22 | 15 | hsa\_circ\_0049224 | , PMID:29606700 |
| 8 | hsa\_circ\_0045272 |  |  |  |

: circRNADisease v2.0.

**Supplementary table 21.** Top 15 predicted results of glioblastoma-related circRNAs based on MKCD.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Rank | CircRNA name | Evidence | Rank | CircRNA name | Evidence |
| 1 | hsa\_circ\_0001801 | , PMID:31456594  , PMID:29967262  , PMID:31493405  , PMID:31823158  , PMID:30738578  , PMID:30470262  , PMID:31802888  , PMID:31599076 | 9 | circMTO1 | , PMID:31456594 |
| 2 | circNT5E | 10 | circ-AKT3 | , PMID:31470874 |
| 3 | circ-PITX1 | 11 | circPVT1 | unconfirmed |
| 4 | hsa\_circ\_0043949 | 12 | circPTN | , PMID:31511040 |
| 5 | hsa\_circ\_0074027 | 13 | hsa\_circ\_101996 | unconfirmed |
| 6 | circMMP9 | 14 | hsa\_circ\_100242 | unconfirmed |
| 7 | circ-Foxo3 | 15 | hsa\_circ\_0003855 | unconfirmed |
| 8 | hsa\_circ\_0001946 |  |  |  |

: circRNADisease v2.0.