**SUPPLEMENTARY DATA**

**Meta-Path-Based Deep Multiple Instance Learning with Heterogeneous Graph Neural Network for Drug-disease Association Prediction**

# Construction of Drug-Disease Network

As for drug-drug similarities and disease-disease similarities, there are quite a few measurements that could be used to calculate the similarities based on chemical structures and disease context information. As the aim of our study is not comparing different similarity measurements, to kepp consistency with other studies, we directly adopted the drug-drug similarity matrixes and disease-disease similarity matrixes of four datasets from previous studies. Specifically, the drug-drug similarity matrixs and disease-disease similarity matrixs for B-dataset, C-datasetr and F-datset, are acquired from [1], where the drug-drug similarities are calculated by: first caculating the binary fingerprint of each drug by the Chemical Development Kit (CDK) [2]. Then, measuring the pairwise similarities based on Tanimoto meansurement; the disease-disease similarities are calculated by: the number of appearance of MeSH (medical subject headings vocabulary) terms of two diseases in the medical descriptions obtained from the OMIM database are measured by MimMiner [3] and obtained as pairwise similarities. The drug-drug similarity matrix and disease-disease similarity matrix for R-dataset are acuiqred from the original study [4], where the drug-drug similarities are calculated by: first calculate the extended-connectivity fingerprint (ECFP) from drug chemical structures. Then, the Tanimoto measurement is used to calculate the pairwise similarities; the disease-disease similarities are calculated by: first convert the MeSH identifiers to hierarchical dirctde acyclie graphs (DAG), which contain ancestral codes for given diseases. Then, the semantic similarities are calculated.

# Experimental Settings and Baseline Methods

As for experimental settings, in the 5-fold cross validation, we collected the prediction results of validation set in each fold. Therefore, after finishing all 5-fold cross validation, all samples in the dataset would get predicted and the calculated metrics on these samples were reported as the results of this 5-fold cross validation.

For the calculation of Pericison, Accuracy, Recall, and Specificity, the thresholds for calculating these metrics is usually set as 0.5 in a balanced dataset (the number of positive samples is approximate to the number of negative samples). However, in the drug-disease association prediction task where the datasets are often much imbalanced, we should set a much lower threshold to response to the exterme label distribution in this scenario. Therefore, we used a strategy of threshold setting proposed in the previous study [5], which has also been adopted to other studies [4, 6, 7]. In this strategy, we evenly set 1000 thresholds from 0 to 1 and calculated metrics (Accuracy, Precision, Recall and Specificity) with different thresholds, respectively. Then, the best performance under the calculation of one threshold would be used and reported. The overall process is like finding a suboptimal Youden index, and we used this strategy in all methods and baselines to guarantee the consistency. Meanwhile, the Pericison, Accuracy, Recall, F1-Score, and Specificity can be formulated as:

1

2

3

5

5

where is the number of true-predicted postitive samples, is the number of true-predicted negative samples, is the number of false-predicted positive samples, is the number of false-predicted negative samples.

As for baseline methods used in our study, they are:

* DDA-SKF [8] is a method using similarity kernel fusion to integrate multiple drug and disease similarities measures, then adopting Laplacian regularized least squares to obtain the prediction association matrix. Due to the lack of required information, we only tested the performance of DDA-SKF with a single chemical similarity kernel.
* SCPMF [9] is a similarity constrained probabilistic matrix factorization method for drug repositioning. It is initially used in drug-virus network, which projects the drug-virus interaction metrix into drug and viruses latent feature matrices, and reconstruct the drug-viruse interactions with a weighted similarity interaction matrix as constraints. As the drug-virus network is similar to drug-disease network, we adopted and reproduced SCPMF for model comparison. It should be noted the reconstructed drug-disease association matrix on “Integrated” dataset predicted by SCPMF contains plenty of NaN values. We filled these NaN values using average prediction scores for pair measurement and comparison.
* NIMCGCN [10] takes GCN as the backbone model and reconstructs the miRNA-disease associations by neural inductive matrix completion. Recently, NIMGCN has been applied in several studies for the model comparisons in drug repositioning [4, 5, 11].
* DRWBNCF [7] is a neighborhood and neighborhood interaction-based neural collaborative filtering approach for drug-disease association prediction.
* REDDA [4] is a heterogeneous graph neural network-based method which utlized multiple biological relations for drug repositioning. It has three attention mechanisms for advanced representation learning of drug and disease nodes.
* PSGCN [11] is a partner-specific method based on graph convolution network, which transforms the DDA prediction, a link prediction task, to a graph classification task. Such partner-specific graph induces more refined local structural features for inferring potential associations.
* HAN [12] is a meta-path-based heterogeneous graph neural network method, which utilizes node-level attention and semantic-level attention for node representation learning. In our experiment, we defined two meta-paths “drug-disease-drug”, and “disease-drug-disease” in the drug-disease networks for the model training.
* MHGNN [13] is a meta-path-based heterogeneous graph neural network method, which initially focuses on drug-target interaction prediction in a biological heterogeneous graph. Specifically, it employs a heterogeneous graph attention network for dual node-level representation learning and a feature integration process for aggregating representations from different meta-path-based graphs. In this study, we defined four meta-paths for drugs: “drug-drug”, “drug-disease-drug”, “drug-disease-disease-drug”, “drug-disease-drug-disease-drug”, and four meta-paths for diseases: “disease-disease”, “disease-drug-disease”, “dieasese-drug-drug-disease”, “disease-drug-disease-drug-disease” for the meta-path-based graph generation. It should be noted that there are two huge information leakage problems in it. That is: (1). It doesn’t delete the tested DDAs in the tested graph. As a result, the model can acquire the neighbor information of tested DDAs during messaging passing procedure in the testing stage. In other words, the model just needs to learn how to “search” these links in the graph. (2). The test set hasn’t been shuffled. The minibatches of the tested DDAs are filled with either all positive samples or all negative samples. As a result, the averages and variances of these minibatches will provide extra information by the batch normalization process used in MHGNN, which cause overconfident prediction. A recentl study has also mentioned similar issues for MHGNN [14]. We revised the information leakage problem by deleting the tested DDAs in the tested graph and random split the tested DDAs for fair performance comparison.

For baseline models, we adopt the reported optimum parameters in the original references and estimate the model performance on our benchmark dataset.

# Performance Comparisons with Baseline Models in the Cross-Validation

The Precision, Accuracy, Recall, and Specificiy results of MilGNet and eight baseline methods on four datasets were listed in **Tabel S1**, and results in “Integrated” dataset were listed in **Table S2**. Due to the high running time, we can only perform MHGNN for a single cross validation on B-dataset and “Integrated” dataset. Meanwhile, we used t-test to calculate the statistical significances of the difference between metric results of MilGNet and those of baseline methods, which were listed in **Table S3-7**. Is should be noted that the results in normal faces mean the significant outperforming of MilGNet compared to the baseline method (*p*-value < 0.05), while the results in *italic faces* mean that the none significant outperforming of MilGNet compared to the baseline method, and those in red faces mean the significant outperforming of baseline method compared to MilGNet (*p*-value < 0.05).

Table S1. Precision, Accuracy, Recall, and Specificity Results in Model Comparision Experiments on Four Datasets

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Dataset | Metric | DDA-SKF | SCPMF | NIMCGCN | DRWBNCF | REDDA | PSGCN | HAN | MHGNN | MilGNet |
| B-dataset |  |  |  |  |  |  |  |  |  |  |
|  | Precision | 0.259±0.005 | 0.468±0.011 | 0.223±0.016 | 0.428±0.013 | 0.444±0.009 | 0.365±0.018 | 0.258±0.030 | 0.136±0.000 | **0.448±0.008** |
|  | Accuracy | 0.790±0.006 | **0.877±0.004** | 0.766±0.017 | 0.865±0.005 | 0.869±0.004 | 0.840±0.009 | 0.782±0.057 | 0.514±0.000 | 0.870±0.003 |
|  | Recall | 0.448±0.013 | 0.559±0.015 | 0.416±0.017 | 0.532±0.020 | 0.556±0.016 | 0.529±0.017 | 0.447±0.067 | **0.605±0.000** | 0.562±0.012 |
|  | Specificity | 0.834±0.009 | **0.918±0.006** | 0.811±0.020 | 0.908±0.008 | 0.910±0.006 | 0.880±0.011 | 0.825±0.073 | 0.502±0.000 | 0.910±0.005 |
| C-dataset |  |  |  |  |  |  |  |  |  |  |
|  | Precision | 0.138±0.012 | 0.531±0.021 | 0.089±0.012 | **0.708±0.029** | 0.591±0.018 | 0.384±0.036 | 0.100±0.007 | 0.055±0.004 | 0.620±0.035 |
|  | Accuracy | 0.982±0.002 | 0.991±0.000 | 0.981±0.003 | **0.993±0.000** | 0.992±0.000 | 0.988±0.001 | 0.974±0.005 | 0.969±0.005 | 0.992±0.000 |
|  | Recall | 0.178±0.024 | 0.414±0.011 | 0.112±0.022 | 0.465±0.016 | 0.481±0.015 | 0.410±0.034 | 0.223±0.038 | 0.142±0.022 | **0.517±0.022** |
|  | Specificity | 0.989±0.003 | 0.997±0.000 | 0.989±0.003 | **0.998±0.000** | 0.997±0.000 | 0.994±0.001 | 0.981±0.005 | 0.977±0.005 | 0.997±0.001 |
| F-dataset |  |  |  |  |  |  |  |  |  |  |
|  | Precision | 0.148±0.015 | 0.440±0.031 | 0.078±0.027 | **0.554±0.030** | 0.533±0.043 | 0.294±0.023 | 0.071±0.006 | 0.057±0.004 | 0.524±0.026 |
|  | Accuracy | 0.981±0.002 | 0.988±0.001 | 0.978±0.007 | **0.990±0.000** | 0.990±0.001 | 0.984±0.001 | 0.959±0.007 | 0.974±0.002 | **0.990±0.000** |
|  | Recall | 0.160±0.014 | 0.356±0.027 | 0.083±0.022 | 0.356±0.008 | 0.399±0.022 | 0.363±0.030 | 0.236±0.033 | 0.098±0.009 | **0.446±0.017** |
|  | Specificity | 0.990±0.002 | 0.995±0.001 | 0.988±0.007 | **0.997±0.000** | 0.996±0.001 | 0.991±0.002 | 0.967±0.007 | 0.983±0.002 | 0.996±0.001 |
| R-dataset |  |  |  |  |  |  |  |  |  |  |
|  | Precision | 0.125±0.004 | **0.500±0.018** | 0.125±0.023 | 0.237±0.029 | 0.402±0.013 | 0.185±0.072 | 0.107±0.006 | 0.068±0.008 | 0.476±0.020 |
|  | Accuracy | 0.984±0.001 | **0.993±0.000** | 0.987±0.002 | 0.991±0.001 | 0.992±0.000 | 0.986±0.003 | 0.983±0.002 | 0.981±0.003 | **0.993±0.000** |
|  | Recall | 0.229±0.011 | 0.382±0.009 | 0.152±0.015 | 0.220±0.027 | 0.362±0.012 | 0.296±0.049 | 0.212±0.029 | 0.138±0.024 | **0.387±0.009** |
|  | Specificity | 0.989±0.001 | **0.997±0.000** | 0.993±0.002 | 0.996±0.001 | 0.996±0.000 | 0.990±0.004 | 0.988±0.002 | 0.987±0.003 | **0.997±0.000** |
| *Avg.* |  |  |  |  |  |  |  |  |  |  |
|  | Precision | 0.168 | 0.472 | 0.129 | 0.482 | 0.493 | 0.307 | 0.134 | 0.079 | **0.517** |
|  | Accuracy | 0.934 | **0.961** | 0.928 | 0.960 | **0.961** | 0.950 | 0.925 | 0.860 | **0.961** |
|  | Recall | 0.254 | 0.433 | 0.191 | 0.393 | 0.450 | 0.400 | 0.280 | 0.246 | **0.478** |
|  | Specificity | 0.951 | **0.975** | 0.945 | **0.975** | **0.975** | 0.964 | 0.940 | 0.862 | **0.975** |
| *Avg. Rank* |  |  |  |  |  |  |  |  |  |  |
|  | Precision | 6.25 | 2.75 | 7.50 | 2.50 | 2.75 | 5.00 | 7.25 | 9.00 | **2.00** |
|  | Accuracy | 6.25 | 2.50 | 6.75 | 1.75 | 2.75 | 5.25 | 8.00 | 8.75 | **1.50** |
|  | Recall | 6.50 | 3.50 | 8.75 | 4.50 | 2.75 | 4.50 | 6.75 | 6.50 | **1.25** |
|  | Specificity | 6.25 | **2.00** | 6.50 | 2.50 | 2.50 | 5.25 | 8.00 | 8.75 | 2.25 |

Table S2. Precision, Accuracy, Recall, and Specificity Results in Model Comparision Experiments on Integrated

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Metric | DDA-SKF | SCPMF | NIMCGCN | DRWBNCF | REDDA | PSGCN | HAN | MHGNN | MilGNet |
| Precision | 0.194±0.002 | 0.487±0.008 | 0.161±0.008 | 0.328±0.010 | 0.448±0.005 | 0.400±0.021 | 0.257±0.031 | 0.116±0.000 | **0.495±0.006** |
| Accuracy | 0.885±0.003 | 0.946±0.001 | 0.888±0.006 | 0.923±0.003 | 0.941±0.001 | 0.934±0.004 | 0.907±0.008 | 0.842±0.000 | **0.947±0.001** |
| Recall | 0.384±0.008 | 0.482±0.007 | 0.271±0.014 | 0.454±0.011 | 0.570±0.009 | 0.516±0.018 | 0.410±0.035 | 0.305±0.000 | **0.579±0.010** |
| Specificity | 0.912±0.003 | **0.972±0.001** | 0.922±0.006 | 0.949±0.003 | 0.961±0.001 | 0.957±0.005 | 0.934±0.008 | 0.872±0.000 | 0.968±0.001 |

Table S3. Statstical Results of AUC, AUPR, F1-Score, Precision, Accuracy, Recall, and Specificity in Model Comparision Experiments on B-dataset

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Model | AUC | AUPR | F1-Score | Precision | Accuracy | Recall | Specificity |
| DDA-SKF | 3.87E-20 | 5.40E-18 | 8.60E-18 | 3.81E-13 | 3.29E-11 | 5.85E-09 | 1.29E-09 |
| SCPMF | 5.97E-04 | 8.52E-08 | 5.04E-07 | 6.60E-03 | 1.06E-02 | *9.79E-01* | *5.23E-02* |
| NIMCGCN | 2.33E-10 | 5.00E-12 | 2.12E-11 | 1.10E-10 | 3.02E-08 | 4.19E-09 | 1.94E-07 |
| DRWBNCF | 2.02E-10 | 2.09E-09 | 5.99E-08 | 3.49E-03 | 1.48E-02 | 9.53E-03 | *3.40E-01* |
| REDDA | 6.84E-08 | *9.44E-02* | 1.10E-03 | *1.88E-01* | *2.49E-01* | *5.50E-01* | *5.46E-01* |
| PSGCN | 5.24E-10 | 3.96E-09 | 7.00E-09 | 3.81E-07 | 2.60E-06 | 5.56E-04 | 2.63E-05 |
| HAN | 1.60E-10 | 1.59E-12 | 4.97E-10 | 1.21E-08 | 8.23E-04 | 4.81E-04 | 4.73E-03 |
| MHGNN | - | - | - | - | - | - | - |

Table S4. Statstical Results of AUC, AUPR, F1-Score, Precision, Accuracy, Recall, and Specificity in Model Comparision Experiments on C-dataset

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Model | AUC | AUPR | F1-Score | Precision | Accuracy | Recall | Specificity |
| DDA-SKF | 2.72E-17 | 9.30E-16 | 1.63E-15 | 9.14E-12 | 1.83E-07 | 2.05E-11 | 6.37E-06 |
| SCPMF | 7.99E-14 | 1.61E-10 | 2.61E-09 | 1.00E-04 | 1.13E-05 | 1.15E-07 | *5.67E-02* |
| NIMCGCN | 1.49E-10 | 3.72E-17 | 1.74E-16 | 1.05E-12 | 2.12E-07 | 4.75E-12 | 8.80E-06 |
| DRWBNCF | 1.03E-13 | *1.77E-01* | *5.00E-01* | 2.17E-04 | 9.66E-04 | 7.57E-05 | 2.20E-04 |
| REDDA | 1.40E-12 | 6.01E-07 | 4.32E-05 | 4.51E-02 | 7.56E-03 | 6.36E-03 | *5.63E-01* |
| PSGCN | 3.24E-07 | 4.72E-08 | 9.71E-09 | 1.61E-07 | 2.75E-07 | 6.36E-06 | 8.23E-06 |
| HAN | 6.86E-15 | 1.11E-15 | 5.42E-15 | 4.79E-12 | 3.52E-07 | 2.77E-09 | 2.41E-06 |
| MHGNN | 2.16E-13 | 2.79E-16 | 2.46E-16 | 1.55E-12 | 5.83E-08 | 1.06E-11 | 3.15E-07 |

Table S5. Statstical Results of AUC, AUPR, F1-Score, Precision, Accuracy, Recall, and Specificity in Model Comparision Experiments on F-dataset

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Model | AUC | AUPR | F1-Score | Precision | Accuracy | Recall | Specificity |
| DDA-SKF | 3.28E-16 | 6.05E-15 | 4.15E-15 | 2.39E-11 | 1.15E-07 | 5.37E-11 | 1.21E-05 |
| SCPMF | 6.14E-10 | 3.68E-10 | 1.49E-09 | 1.78E-05 | 1.03E-04 | 6.92E-06 | *1.17E-01* |
| NIMCGCN | 1.02E-09 | 2.26E-16 | 5.72E-14 | 1.63E-11 | 3.47E-04 | 1.59E-11 | 5.48E-03 |
| DRWBNCF | 2.96E-12 | 9.86E-09 | 1.33E-07 | 2.62E-02 | *6.31E-02* | 4.30E-07 | 1.39E-04 |
| REDDA | 1.77E-11 | 5.49E-06 | 5.04E-05 | *6.22E-01* | *8.86E-01* | 2.39E-03 | *1.68E-01* |
| PSGCN | 1.11E-05 | 1.19E-10 | 2.86E-10 | 1.00E-08 | 3.90E-07 | 1.15E-05 | 5.78E-06 |
| HAN | 5.71E-14 | 2.22E-16 | 1.85E-16 | 5.64E-13 | 1.90E-07 | 3.05E-08 | 5.57E-07 |
| MHGNN | 1.46E-16 | 1.57E-16 | 5.01E-17 | 2.90E-13 | 1.01E-09 | 1.85E-13 | 1.27E-08 |

Table S6. Statstical Results of AUC, AUPR, F1-Score, Precision, Accuracy, Recall, and Specificity in Model Comparision Experiments on R-dataset

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Model | AUC | AUPR | F1-Score | Precision | Accuracy | Recall | Specificity |
| DDA-SKF | 3.00E-13 | 1.23E-14 | 8.68E-16 | 6.22E-13 | 2.36E-11 | 3.36E-10 | 2.30E-10 |
| SCPMF | 1.17E-02 | 7.56E-02 | 1.82E-02 | 2.06E-02 | 2.57E-02 | *3.24E-01* | 3.99E-02 |
| NIMCGCN | 3.17E-08 | 9.37E-13 | 3.51E-12 | 8.59E-12 | 1.82E-06 | 4.45E-12 | 2.06E-05 |
| DRWBNCF | 3.36E-08 | 4.47E-11 | 3.04E-10 | 3.71E-09 | 1.14E-05 | 3.65E-08 | 4.36E-04 |
| REDDA | 7.12E-07 | 4.77E-08 | 6.09E-07 | 4.48E-06 | 3.29E-06 | 2.38E-03 | 9.77E-05 |
| PSGCN | 1.21E-06 | 9.10E-08 | 8.51E-08 | 1.50E-06 | 7.50E-05 | 3.54E-04 | 2.34E-04 |
| HAN | 1.35E-12 | 1.28E-15 | 1.64E-15 | 3.02E-13 | 4.89E-08 | 9.21E-09 | 2.85E-07 |
| MHGNN | 9.38E-12 | 3.85E-15 | 7.33E-15 | 1.90E-12 | 1.14E-06 | 5.82E-11 | 5.53E-06 |

Table S7. Statstical Results of AUC, AUPR, F1-Score, Precision, Accuracy, Recall, and Specificity in Model Comparision Experiments on Integrated

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Model | AUC | AUPR | F1-Score | Precision | Accuracy | Recall | Specificity |
| DDA-SKF | 4.40E-14 | 5.15E-10 | 1.63E-11 | 2.80E-08 | 1.37E-07 | 8.72E-08 | 3.15E-07 |
| SCPMF | 5.98E-10 | 4.75E-07 | 3.19E-08 | *8.15E-02* | *1.20E-01* | 7.59E-05 | 1.56E-02 |
| NIMCGCN | 1.06E-06 | 5.28E-09 | 6.64E-08 | 2.41E-08 | 3.93E-05 | 4.49E-06 | 1.88E-04 |
| DRWBNCF | 5.95E-07 | 1.21E-06 | 5.92E-07 | 3.00E-05 | 1.74E-04 | 5.41E-05 | 1.29E-03 |
| REDDA | 4.97E-07 | 1.07E-05 | 4.81E-07 | 6.48E-04 | 8.88E-04 | *1.91E-01* | 7.27E-03 |
| PSGCN | 4.79E-05 | 5.40E-04 | 3.49E-04 | 8.49E-04 | 3.18E-03 | 1.60E-04 | 9.04E-03 |
| HAN | 1.34E-03 | 4.83E-05 | 6.40E-05 | 6.03E-05 | 2.25E-04 | 3.39E-04 | 5.14E-04 |
| MHGNN | - | - | - | - | - | - | - |

# Ablation Study

We designed six simplified variants to desmonstrate the reasonability of the overall framework of MilGNet, which are listed as follows:

* MilGNet-w/o LayerAttn: The MilGNet model without layer attention machansim in HGNN-based node encoder.
* MilGNet-w/o Sum: The MilGNet model, which is equipped with Sum calculation as a substitute of BiTrans.
* MilGNet-w/o Mean: The MilGNet model, which is equipped with Mean calculation as a substitute of BiTrans.
* MilGNet-w/o Linear: The MilGNet model, which is equipped with Linear neural network as a substitute of BiTrans.
* MilGNet-w/o AttbAgg: The MilGNet model without the attention-based aggregation in the aggregation of instance embeddings, but with a Mean calculation as substitute.
* MilGNet-w/o InsPred: The MilGNet model without instance predictor in multi-scale interpretable predictor.

# Model Hyper-parameters

As described in Parameter Analysis section, we executed a sequential hyper-parameter search to investigate optimal hyper-parameter settings of MilGNet on five datasets. The optimal hyper-parameter settings were listed in **Table S8**.

Table S8. Optimal Hyper-parameter Settings of MilGNet for Five Datasets

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Hyper-parameters | Dataset | | | | |
| B-dataset | C-dataset | F-dataset | R-dataset | Integrated |
| epoch | 1000 | 2000 | 2000 | 2000 | 2000 |
| learning rate | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| batch size | 2048 | 1024 | 1024 | 4096 | 2048 |
| dropout rate | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 |
| number of HGNN layer | 2 | 1 | 1 | 2 | 2 |
| dimension of hidden features | 32 | 128 | 64 | 128 | 64 |
| *topk* binarization | 5 | 5 | 10 | 5 | 5 |

# Model Interpretability

As described in model interpretability section in the manuscript, the top3 meta-paths mentioned three diseases: Mismatch Repair Cancer Syndrome, Familial Adenomatous Polyposis, and Colorectal Cancer. To enhance the reasonability of our predicted meta-paths, we plotted the phenotype-gene association linear graph of these diseases to find out common associated phenotypes and genes based on OMIM, which were shown in **Fig. S1-3**.



Figure S1. Phenotype-Gene Association Linear Graph for Mismatch Repair Cancer Syndrome



Figure S2. Phenotype-Gene Association Linear Graph for Familial Adenomatous Polyposis

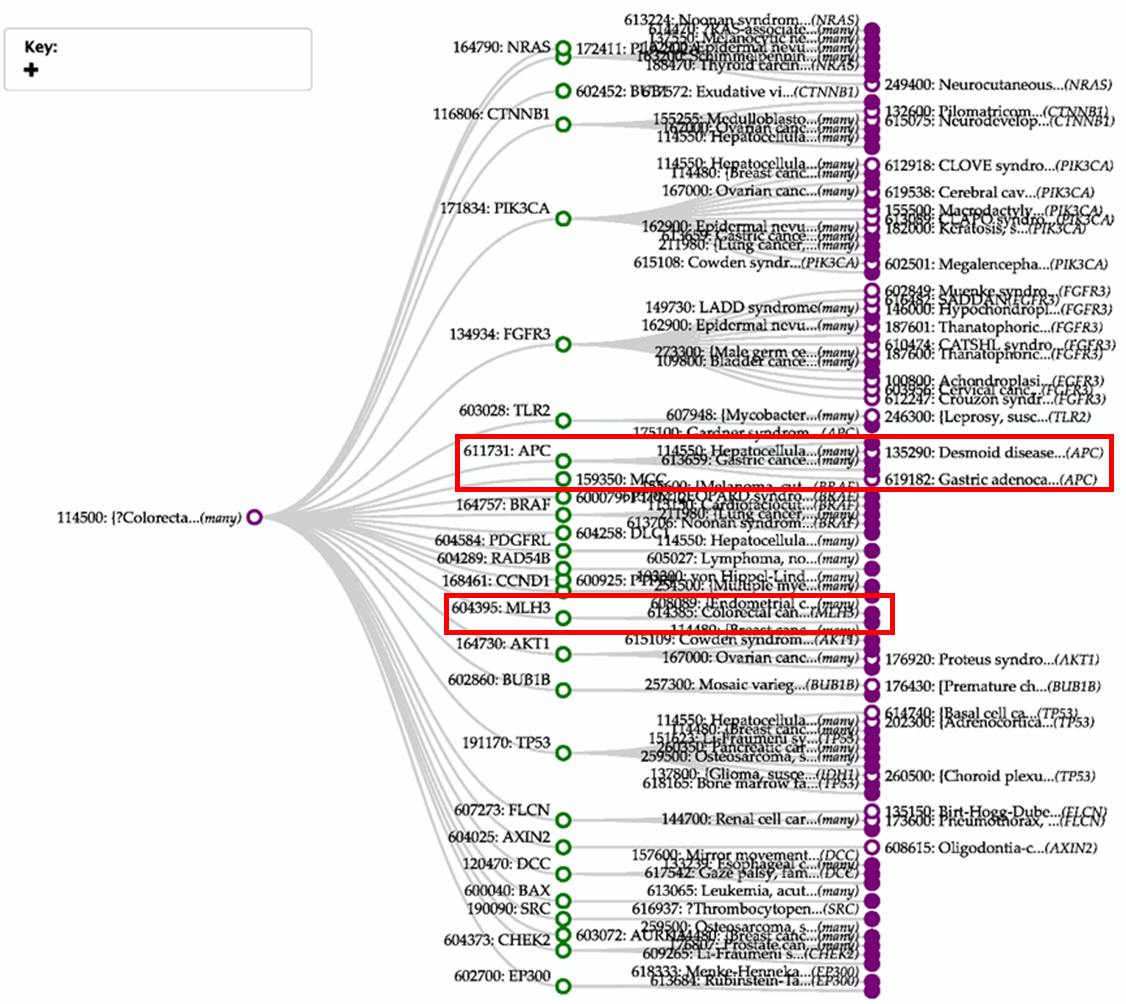


Figure S3. Phenotype-Gene Association Linear Graph for Colorectal Cancer

# Reference

[1] M. Yang, H. Luo, Y. Li, and J. Wang, "Drug repositioning based on bounded nuclear norm regularization," (in eng), *Bioinformatics (Oxford, England),* vol. 35, no. 14, pp. i455-i463, Jul 15 2019.

[2] C. Steinbeck, Y. Han, S. Kuhn, O. Horlacher, E. Luttmann, and E. Willighagen, "The Chemistry Development Kit (CDK): an open-source Java library for Chemo- and Bioinformatics," (in eng), *Journal of chemical information and computer sciences,* vol. 43, no. 2, pp. 493-500, Mar-Apr 2003.

[3] M. A. van Driel, J. Bruggeman, G. Vriend, H. G. Brunner, and J. A. Leunissen, "A text-mining analysis of the human phenome," (in eng), *European journal of human genetics : EJHG,* vol. 14, no. 5, pp. 535-42, May 2006.

[4] Y. Gu, S. Zheng, Q. Yin, R. Jiang, and J. Li, "REDDA: Integrating multiple biological relations to heterogeneous graph neural network for drug-disease association prediction," (in eng), *Computers in biology and medicine,* vol. 150, p. 106127, Sep 22 2022.

[5] Z. Yu, F. Huang, X. Zhao, W. Xiao, and W. Zhang, "Predicting drug-disease associations through layer attention graph convolutional network," (in eng), *Briefings in bioinformatics,* vol. 22, no. 4, Jul 20 2021.

[6] L. Cai *et al.*, "Drug repositioning based on the heterogeneous information fusion graph convolutional network," (in eng), *Briefings in bioinformatics,* vol. 22, no. 6, Nov 5 2021.

[7] Y. Meng, C. Lu, M. Jin, J. Xu, X. Zeng, and J. Yang, "A weighted bilinear neural collaborative filtering approach for drug repositioning," (in eng), *Briefings in bioinformatics,* vol. 23, no. 2, Mar 10 2022.

[8] C.-Q. Gao, Y.-K. Zhou, X.-H. Xin, H. Min, and P.-F. Du, "DDA-SKF: Predicting Drug-Disease Associations Using Similarity Kernel Fusion," (in eng), *Front Pharmacol,* vol. 12, pp. 784171-784171, 2022.

[9] Y. Meng, M. Jin, X. Tang, and J. Xu, "Drug repositioning based on similarity constrained probabilistic matrix factorization: COVID-19 as a case study," (in eng), *Applied soft computing,* vol. 103, p. 107135, May 2021.

[10] J. Li, S. Zhang, T. Liu, C. Ning, Z. Zhang, and W. Zhou, "Neural inductive matrix completion with graph convolutional networks for miRNA-disease association prediction," (in eng), *Bioinformatics (Oxford, England),* vol. 36, no. 8, pp. 2538-2546, Apr 15 2020.

[11] X. Sun, B. Wang, J. Zhang, and M. Li, "Partner-Specific Drug Repositioning Approach Based on Graph Convolutional Network," (in eng), *IEEE journal of biomedical and health informatics,* vol. 26, no. 11, pp. 5757-5765, Nov 2022.

[12] X. Wang *et al.*, "Heterogeneous graph attention network," in *The world wide web conference*, 2019, pp. 2022-2032.

[13] M. Li, X. Cai, S. Xu, and H. Ji, "Metapath-aggregated heterogeneous graph neural network for drug-target interaction prediction," (in eng), *Briefings in bioinformatics,* vol. 24, no. 1, Jan 19 2023.

[14] Q. Lv *et al.*, "Are we really making much progress? revisiting, benchmarking and refining heterogeneous graph neural networks," in *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*, 2021, pp. 1150-1160.