Predicting miRNA-disease association through combining miRNA function and network topological similarities based on MINE

Example about extending miRNA function module based on EC

Examples of the identified miRNA functional modules by our proposed method
Predicting miRNA-disease association through combining miRNA function and network topological similarities based on MINE

Buwen Cao,1,2,3,* Renfa Li,2,* Sainan Xiao,1,2 Shuguang Deng,1 Xiangjun Zhou,1 and Lang Zhou1

SUMMARY

Predicting associations between microRNAs (miRNAs) and diseases from the viewpoint of function modules has become increasingly popular. However, existing methods obtained the relations between diseases and miRNAs only through the construction of similarity networks and neglected the complex network characteristic. In this paper, a new method named combining miRNA function similarities and network topology similarities based on module identification in networks (ComSim-MINE) was developed. Combined similarity is calculated from the harmonic mean between miRNA function similarities and network topology similarities. Experimental results showed that ComSim-MINE can compete with several state-of-the-art weighted function module algorithms, such as ClusterONE, MCODE, NEMO, and SPICi, and achieved the satisfactory results in terms of the composite score of F-measure, sensitivity, and accuracy based on the generated miRNA function interaction network. From the analysis of case studies, some new findings obtained from our proposed method provide clinicians new clues for epidemic diseases, such as COVID-19.

INTRODUCTION

The identification of associations between diseases and miRNAs has attracted much attention. Accumulating evidence has indicated that miRNAs play key roles in biological processes and complex diseases (Alvarez-Garcia and Miska, 2005; Lynam-Lennon et al., 2009; Wu et al., 2021). Therefore, predicting disease-associated miRNAs will be beneficial for researching the mechanisms of pathogenicity and promoting the diagnosis and treatment of human diseases.

Numerous computational methods have been proposed to detect potential miRNA–disease associations based on the assumption that functionally similar miRNAs are likely to be associated with similar diseases, and vice versa (Luo et al., 2017a, 2017b; Zou et al., 2016). The majority of methods that identify the associations between diseases and miRNAs are consisted of three aspects: computing similarity, constructing similarity networks, and detecting the associations between diseases and miRNAs based on algorithm models. Pan et al. (Pan and Shen, 2020) proposed a semi-supervised multi-label graph convolutional network to infer disease-associated miRNAs on an interaction network between protein-coding genes and miRNAs using disease–protein-coding gene associations. Chen et al. (Chen et al., 2021b) proposed a new computational model of neighborhood constraint matrix completion to predict potential miRNA–disease associations based on known miRNA–disease associations and integrated disease (miRNA) similarities. Different from previous studies, a novel edge perturbation method was proposed for predicting potential miRNA–disease associations (Dong et al., 2020). In this method, the extracted features (each edge of a graph by structural Hamiltonian information) are used to train a multilayer perception model to predict the candidatedisease–miRNA associations. Li et al. (Li et al., 2021) developed a novel network projection-based dual random walk with restart to predict potential disease-related miRNAs. It can also be used to predict isolated diseases and new miRNAs. Xuan et al. (Xuan et al., 2020) proposed a non-negative matrix factorization-based method for predicting disease miRNAs in heterogeneous networks with node attributes. Similarly, a computational model of matrix decomposition was developed to discover miRNA–disease association by integrating the predicted association probability obtained from matrix decomposition through sparse learning method, the miRNA functional similarity, the disease semantic similarity, and the Gaussian interaction profile kernel similarity for diseases and miRNAs into a heterogeneous...
network (Chen et al., 2018d). A matrix factorization model was established to predict lncRNA–miRNA interactions by logistic matrix factorization with neighborhood regularized, and detected potential lncRNA–miRNA associations (Liu et al., 2020). Xiao et al. (Xiao et al., 2021) presented a novel computational method with adaptive multisource multi-view latent feature learning to infer potential disease-associated miRNAs; the experimental results showed the effectiveness of the proposed method. Chen et al. (Chen et al. (2021a) developed a novel model of deep-belief network for miRNA–disease association prediction. Zhu et al. (Zhu et al., 2021) raised a novel model of Bayesian ranking for miRNA–disease association prediction by improving Bayesian personalized ranking, leave-one-out cross-validation and area under the receiver operating characteristic curve validated the performance of the proposed method. Chen et al. (Chen et al. (2020) presented a neoteric Bayesian model that combines kernel-based nonlinear dimensionality reduction, matrix factorization, and binary classification; the main idea of which is to project miRNAs and diseases into a unified subspace and estimate the association network in that subspace. Peng et al. (Peng et al. (2020) developed a novel prediction model named ensemble of kernel ridge regression-based miRNA–disease association prediction and evaluated the model by global and local leave-one-out cross-validation. Zhao et al. (Zhao et al. (2020) presented a computational model named adaptive boosting for miRNA–disease association prediction, which can balance positive and negative samples by performing random sampling based on K-means clustering on negative samples to predict the potential associations between diseases and miRNAs. However, none of these methods obtained excellent performance in the identification of latent disease–miRNA relationship, and most of them relied on top-ranked association model.

Other computational models about disease–miRNA association identification have been proposed. The first model is derived from mathematical model, such as laplacian regularized sparse subspace learning (Chen and Huang, 2017), decision tree (Chen et al., 2019b), stacked autoencoder (Wang et al., 2022), and logistic model tree (Wang et al., 2019). Another model is originated from the decomposition based on the network topology (Chen et al., 2018c; Liu et al., 2022a; You et al., 2017; Zhang et al., 2021a). Besides, the random forest algorithm and its variant are developed to identify the miRNA–disease associations (Chen et al., 2018a; Liu et al., 2022b).

Recent studies showed that miRNA function modules or clusters play a key role in many diseases (Cheng et al., 2014; Kandettu et al., 2020; Yoshida et al., 2021). Many computational models infer disease–miRNA associations from the viewpoint of function module (cluster) identification. Nalluri et al. (Nalluri et al. (2017) designed a novel computational pipeline to predict the common core sets of miRNA–miRNA interactions for different diseases using network inference algorithms on miRNA–disease expression profiles. Luo et al. (Luo et al. (2019) presented a novel cluster-based computational method to identify miRNA regulatory modules; miRNAs in each module are expected to present cooperative mechanisms in regulating their target mRNAs. We developed an improved K-means algorithm to detect miRNA function module based on the miRNA function similarity network derived from a disease similarity network and a known miRNA–disease relationship network (Cao et al., 2021). A framework named MDA-TOEPGA was developed to identify miRNA–disease association based on a two-objective evolutionary programming genetic algorithm (Buwen Cao et al., 2022). The function similarity network is regarded as an unweighted network in the stage of identifying miRNA function modules (Buwen Cao et al., 2022; Cao et al., 2021). However, the function similarity network has the characteristics of a complex network, which can provide researchers some remarkable ideas for identifying miRNA function modules.

Inspired by (Cao et al., 2022; Rhrissorrakrai and Gunsalus, 2011), a method called combination between miRNA function similarities and network topology similarities based on module identification algorithm in networks (ComSim-MINE) was developed to predict miRNA–disease association from the viewpoint of function module. The innovation of this study is that the miRNA function interaction network is constructed from the harmonic mean between miRNA function similarities and network topology similarities. It effectively reduces the impact of combination parameters. Compared with several state-of-the-art weighted function module algorithms, such as ClusterONE, MCODE, NEMO, and SPICi, ComSim-MINE achieved the satisfactory results in terms of the composite score of F-measure, sensitivity, and accuracy based on the generated miRNA function interaction network. Case studies showed that the new findings will provide clinicians with new clues for epidemic diseases, including COVID-19. Figure 1 presents the overall flowchart of our proposed method.

The main contribution of this paper is as follows: first, miRNA similarities are calculated from the harmonic mean of miRNA function similarity and network topological characteristics through the analysis of the
corresponding dataset. Second, extending coefficient (EC) was developed to refine edges with low similarities during the procedure of the ComSim-MINE algorithm to further improve the identification accuracy of associations between diseases and miRNAs. Please see the revised parts marked in yellow.

RESULTS

Similar to existing module identification algorithms (Buwen Cao et al., 2022; Cao et al., 2016, 2021; Jiang and Singh, 2010; Nepusz et al., 2012; Rhrissorrakrai and Gunsalus, 2011; Rivera et al., 2010), the following metrics were used to assess the performance of our proposed method: precision, recall, F-measure, sensitivity, and accuracy.

Impact of EC on prediction performance

In the miRNA function module identification process, EC was proposed to filter lower similarities so as to remarkably improve the clustering process and performance of the ComSim-MINE algorithm. EC values of 0–0.9 with a 0.1 increment were used to achieve the optimal result. The detailed experimental results of F-measure with different EC thresholds are shown in Figure 2.

As shown in Figure 2, the F-measure of the detected miRNA function modules increased slightly with the increasing EC threshold when \( EC \leq 0.5 \). The F-measure reached the maximum value of 0.421 when \( EC = 0.6, 0.7 \). The F-measure declined sharply with the increasing of EC, and an F-measure of 0.006 was obtained when \( EC = 0.9 \). Therefore, the EC value was set to 0.6 in this study.

Analysis of combined similarities

The accuracy of the identified miRNA function modules can be remarkably improved by considering network similarities [43]. The impact of different similarity tactics imposed on the MINE algorithm is investigated to show the performance of the ComSim-MINE algorithm (See Table S1 MINE comparison). Table 1 presents the comparison results using the MINE algorithm on the miRNA function interaction network with different similarities.

Table 1 presented that the ComSim-MINE algorithm outperforms the traditional MINE algorithm with different similarities or without similarities. Notably, although other similarities (e.g., network topological similarity and miRNA function similarity) can effectively improve the performance of the MINE algorithm, our proposed method shows satisfied performance on the miRNA function interaction network. This finding could be attributable to the introduction of the EC, which facilitates the filtering of miRNAs with lower similarities during identification and accelerates the clustering process. Interestingly, the NetSim-MINE, MiSim-MINE, and AvgSim-MINE methods achieved the same experimental results on the metrics of precision, recall, and F-measure. The possible reason is that the construction of the miRNA function interaction network is derived from the computation of miRNA function similarity and miRNA network topology, which is inherent feature in the miRNA function similarity network. Notably, MklSim-MINE obtained

---

Figure 1. Overall flowchart of ComSim-MINE
unsatisfactory experimental results, the main reason is that the data source is not enough abundant. That is, the multiple kernel learning and the similarity network fusion are suit for multiple data source or heterogeneous data sources to compute similarity. SnfSim-MINE is less than that of MKLsim-mine, which failed to obtain effective miRNA functional modules.

Moreover, the combination of network topological similarity and miRNA function similarity imposed on the MINE algorithm can effectively improve the identification of miRNA function modules.

Comparison with other methods
Four state-of-the-art weighted module clustering algorithms, namely, ClusterONE (Nepusz et al., 2012), MCODE (Bader and Hogue, 2003), NEMO (Rivera et al., 2010), and SPICi (Jiang and Singh, 2010), which can detect the function module in weighted biological networks [39], were compared with the proposed method. Parameters required in each method are set as suggested by their authors. The results are listed in Table 2. The highest value of each dataset is presented in bold. The detailed comparison is related to “Table S2 MINE with other algorithms”.

Table 2 provides the detailed experimental results of ComSim-MINE and the other methods using the benchmark dataset. The proposed method identifies 453 miRNA function modules, provided more valuable clues for inferring the latent associations between diseases and miRNAs, and achieved the highest precision and accuracy of 1 and 0.267, respectively. ClusterONE predicts 21 miRNA function modules and achieved the highest sensitivity of 0.919. NEMO identify 119 miRNA function modules and achieved the maximum precision, recall, and F-measure of 1, 0.313, and 0.477, respectively. ComSim-MINE achieved the second highest values of recall, F-measure, and sensitivity, which are 0.267, 0.421, and 0.886, respectively. Interestingly, ComSim-MINE achieved the satisfactory results in terms of the composite score of F-measure, sensitivity, and accuracy (1.574), which is 4.6% higher than the second highest value (obtained by NEMO). Figure 3 presents the comparison results. The detailed results about extending coefficient are related to “Table S3 Extending coefficient”.

DISCUSSION
ComSim-MINE exhibits satisfied performance when tested on a miRNA function interaction network. Several case studies were discussed for potential mapping with known associations between diseases and miRNAs reported in literatures or databases to further show the performance of our proposed method for the miRNA–disease association prediction. The corresponding analysis is presented as follows.
The first case study is module #451 (hsa-mir-599, hsa-mir-572, hsa-mir-614, and hsa-mir-648, as shown in Figure 4A) in multiple sclerosis. Existing studies showed that 22 miRNAs are related to multiple sclerosis. This module demonstrated satisfaction prediction performance for some diseases. For the other four algorithms, ClusterONE and MCODE exactly identify the module #451, SPICi finds three miRNAs (hsa-mir-599, hsa-mir-614, and hsa-mir-648) associated with the module #451, but NEMO does not find the module or miRNAs associated with it.

The second case study is module #161, which includes seven miRNAs (hsa-mir-561, hsa-mir-1179, hsa-mir-1183, hsa-mir-1275, hsa-mir-1909, hsa-mir-552, and hsa-mir-567, as shown in Figure 4B). Among which, three miRNAs are associated with rectal neoplasms; whereas hsa-mir-1179, hsa-mir-1275, and hsa-mir-567 are related to colorectal neoplasms; and hsa-mir-552 is associated with colonic neoplasms. The results show that the miRNA function module #161 is related to three diseases. Clinicians can find commonalities in the treatment of the three diseases using this module; therefore, the module provides valuable references for the treatment of different diseases. It also helps in the construction of clinical treatment networks for human diseases (Mei et al., 2021). Similarly, ClusterONE and NEMO detect all miRNAs in the module #161, in which only hsa-mir-561 is associated with it identified by MCODE and SPICi. It shows that our proposed algorithm has a significant advantage in discovering the potential association of multiple diseases.

ComSim-MINE predicts a miRNA unit that consists of hsa-mir-142 and hsa-mir-155. The unit appears 22 times in the known disease–miRNA association dataset (Wang et al., 2010), accounting for 6.74% of all disease–miRNA associations, and occurred 137 times in the identified modules, accounting for 30.24% of the total modules. The result demonstrates that some small miRNA units play an important role in exploring the potential associations between diseases. Unfortunately, the miRNA unit appeared only once in miRNA function modules found by ClusterONE, MCODE, and SPICi, accounting for 0 of the total modules detected by NEMO.

| Methods        | Number | Precision | Recall | F-measure | Sensitivity | Accuracy |
|----------------|--------|-----------|--------|-----------|-------------|----------|
| ClusterONE     | 21     | 1         | 0.071  | 0.132     | 0.919       | 0.263    |
| MCODE          | 18     | 1         | 0.089  | 0.163     | 0.808       | 0.217    |
| NEMO           | 119    | 1         | 0.313  | 0.477     | 0.861       | 0.19     |
| SPICi          | 26     | 1         | 0.104  | 0.189     | 0.819       | 0.248    |
| ComSim-MINE    | 453    | 1         | 0.267  | 0.421     | 0.886       | 0.267    |

Notes: The highest value is shown in bold. The “Number” refers to the number of the identified miRNA function modules. The “MINE” means the original MINE algorithm, which parameters are set as recommended by authors. The “NetSim-MINE”, “MiSim-MINE”, “AvgSim-MINE”, “MklSim-MINE”, “SnfSim-MINE”, “Comb-MINE” refer to the modified MINE algorithm with different similarities from network topological similarity, miRNA function similarity, average similarity between function similarity and network similarity, similarity calculated from the multiple kernel learning (Wang et al., 2021b), similarity calculated from the similarity network fusion (Wang et al., 2014) and combined similarity between network topological similarity and miRNA function similarity, respectively. “-” means SnfSim-MINE failed to obtain effective miRNA functional modules. The parameters of other modified MINE are set as follows: \(vmp=0.1, msp=0.1, EC=0.6\) in miRNA function interaction network.
Acute lung injury is related to only one miRNA (hsa-mir-16) based on the miRNA–disease associations download from Ref. (Wang et al., 2010). Hsa-mir-16 is also related to 49 diseases, including autoimmune diseases and breast neoplasms. Notably, ComSim-MINE identified 163 miRNA function modules associating hsa-mir-16, which is 1, 1, 14 times, and 1 in the function modules found by ClusterONE, MCODE, NEMO, and SPICi, respectively. The 163 miRNA functional modules are thought to overlap with each other and the diagnosis and treatment of acute lung injury can be sought from the treatment options for other diseases. The results provide valuable ideas for the diagnosis and treatment of epidemic cases. For example, coronavirus disease 2019 (COVID-19) is related to hsa-mir-124 (Zhang et al., 2021b), which is
associated with 38 diseases (http://www.cuilab.cn/hmdd), such as Alzheimer, astrocytoma, and non-small-cell lung carcinoma. Therefore, the associations between hsa-mir-124 and other diseases may be a feasible strategy for determining the prevention and cure of COVID-19.

Identifying potential miRNA–disease associations can help provide valuable references for clinical diagnosis, treatment, and pharmacy. Existing methods for predicting disease-related miRNAs rely on miRNA similarity networks and neglect the important characteristics of complex networks. ComSim-MINE is developed to predict miRNA–disease associations from the viewpoint of function modules to MINE as many associations between diseases and miRNAs as possible. The miRNA function interaction network was constructed from the harmonic mean between miRNA function similarities and network topology similarities to compensate the shortcomings of using a single feature. The experimental results showed that the proposed method effectively improved the identification of the miRNA function modules. Case studies further demonstrated its effectiveness in disease–miRNA association prediction. Interestingly, the associations between hsa-mir-124 and other diseases provide remarkable references for the prevention and cure of COVID-19.

Although ComSim-MINE was better than other classical algorithms in terms of the composite score of F-measure, sensitivity, and accuracy, its overall performance is not very ideal. The main reason is that some known diseases are associated with few miRNAs during the benchmark, and many new associations between diseases and miRNAs cannot be matched with known associations. This barrier can be effectively improved with the development of biotechnology and computing methods, such as artificial intelligence. ComSim-MINE is also expected to become an effective and accurate prediction method for disease–miRNA associations.

Limitations of the study
A limitation of our study is that the combination of network topological similarity and miRNA function similarity imposed on the MINE algorithm. Although it can improve the identification of miRNA function modules, it does not consider multiple sources, only miRNA function similarity and network topological feature. Unfortunately, in this study, for those methods for multiple sources, such as MklSim-MINE, SnfSim-MINE, which are usually employed to seek for similarity in multiple sources, did not achieve perfect experimental results. In the next work, we will improve the identification accuracy of the association between disease and miRNA with the aid of multiple data sources.

STAR METHODS
Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- METHOD DETAILS
  - MiRNA function similarity
  - MiRNA network similarity
  - MiRNA combination similarity
  - Experimental models
  - ComSim-MINE algorithm

SUPPLEMENTAL INFORMATION
Supplemental information can be found online at https://doi.org/10.1016/j.isci.2022.105299.

ACKNOWLEDGMENTS
This work was supported in part by the National Natural Science Foundation of China under Grant 61873089, 62032007 and the Key Project of the Education Department of Hunan Province under Grant 20A087, the Innovation Platform Open Fund Project of Hunan Provincial Education Department under Grant 20K025.
AUTHOR CONTRIBUTIONS
Conceptualization, B.C., R.L., and S.D.; Data curation, S.X.; Funding acquisition, B.C., R.L., and S.X.; Methodology, B.C.; Resources, X.Z. and L.Z.; Validation, X.Z. and L.Z.

DECLARATION OF INTERESTS
The authors declare no competing of interests.
Luo, J., Xiao, Q., Liang, C., and Ding, P. (2017b). Predicting microRNA-disease associations using kronecker regularized least squares based on heterogeneous omics data. IEEE Access 5, 2503–2513. https://doi.org/10.1109/Access.2017.2672500.

Lynam-Lennon, N., Maher, S.G., and Reynolds, J.V. (2009). The roles of microRNA in cancer and apoptosis. Biol. Rev. Camb. Philos. Soc. 84, 55–71. https://doi.org/10.1111/j.1469-185X.2008.00061.x.

Mei, H., Jia, R.F., Qiao, G.Z., Lin, Z.Q., and Ma, S.G. (2021). Human disease clinical treatment network for the elderly: analysis of the medicare inpatient length of stay and readmission data. Biometrics. https://doi.org/10.1111/biom.13549.

Nalluri, J.J., Barh, D., Azevedo, V., and Ghosh, P. (2017). miRig: a consensus-based network inference methodology to identify pan-cancer miRNA-miRNA interaction signatures. Sci Rep-Uk 7, 96384.

Nepusz, T., Yu, H., and Paccanaro, A. (2012). Detecting overlapping protein complexes in protein-protein interaction networks. Nat. Methods 9, 471–472. https://doi.org/10.1038/nmeth.1938.

Pan, X.Y., and Shen, H.B. (2020). Scoring disease-microRNA associations by integrating disease hierarchy into graph convolutional networks. Pattern Recog. 105, 107385.

Peng, L.H., Zhou, L.Q., Chen, X., and Piao, X. (2020). A computational study of potential microRNA-disease association inference based on stacked autoencoder. Brief Bioinform. 22, bbab286.

Wang, B., Mezlini, A.M., Demir, F., Fumie, M., Tu, Z., Brudno, M., Halbe-Kains, B., and Goldenberg, A. (2014). Similarity network fusion for aggregating data types on a genomic scale. Nat. Methods 17, 333–337. https://doi.org/10.1038/nmeth.2810.

Wang, C.C., Han, C.D., Zhao, Q., and Chen, X. (2021a). Circular RNAs and complex diseases from experimental results to computational models. Brief Bioinform. 22.

Wang, C.C., Han, C.D., Zhao, Q., and Chen, X. (2021b). Circular RNAs and complex diseases: from experimental results to computational models. Brief Bioinform. 22.

Xiao, Q., Luo, J.W., Liang, C., Cai, J., and Ding, P.J. (2018). A graph regularized non-negative matrix factorization method for identifying microRNA-disease associations. Bioinformatics 34, 239–248.

Xiao, Q., Zhang, N., Luo, J.W., Dai, J.H., and Tang, X.W. (2021). Adaptive multi-source multi-view latent feature learning for inferring potential disease-associated miRNAs. Brief Bioinform. 22, 2043–2057.

Xuan, P., Shen, T.H., Wang, X., Zhang, T.G., and Zhang, W.X. (2020). Inferring disease-associated microRNAs in heterogeneous networks with node attributes. IEEE Access. Biol. Comput. Biol. 17, 1019–1031.

Yoshida, K., Yokoi, A., Yamamoto, Y., and Kajiyama, H. (2021). ChiQ27.3 miRNA cluster functions in cancer development. J Exp Clin Canc Res-40, 105261. https://doi.org/10.1016/J.Knosys.2019.

You, Z.H., Huang, Z.A., Zhu, Z.X., Yan, G.Y., Li, Z.W., Wen, Z.K., and Chen, X. (2017). PBMDA: A novel and effective path-based computational model for miRNA-disease association prediction. PLoS Comput. Biol. 13.

Zhang, S., Amahong, K., Sun, X.N., Lian, X.C., Liu, J., Sun, H.C., Lou, Y., Zhu, F., and Qiu, Y.Q. (2021b). The miRNA: a small but powerful RNA for COVID-19. Brief Bioinform. 22, 1137–1149.

Zhang, L., Yang, P.Y., Feng, H.W., Zhao, Q., and Liu, H.S. (2021a). Using Network Distance Analysis to Predict lncRNA-miRNA Interactions. Interdiscip. Sci. 13, 535–545.

Zhao, Y., Chen, X., and Yin, J. (2020). Adaptive boosting-based computational model for predicting potential miRNA-disease associations (vol 47, pg 353, 2019). Bioinformatics 36, 330. https://doi.org/10.1371/journal.pcbi.1006418.

Zhu, C.C., Wang, C.C., Zhao, Y., Zuo, M.C., and Chen, X. (2021). Identification of miRNA-disease associations via multiple information integration with Bayesian ranking. Brief Bioinform. 22.
STAR METHODS

KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---------------------|--------|------------|
| Deposited data      |        |            |
| MiRNA function similarity data | This paper | http://www.cuilab.cn/files/images/cuilab/misim.zip |
| hsa-mir-124         | This paper | http://www.cuilab.cn/hmdd |
| Software and algorithms |      |            |
| MINE                | This paper | http://www.cytoscape.org |
| Other               |        |            |
| Materials           | This paper | N/A |
| Data and code       | This paper | N/A |

RESOURCE AVAILABILITY

Lead contact
The raw data, analytic methods, and study materials will be publicly available as online-only Supplement information. Study materials will be provided after a reasonable request. Inquiries can be directed to the lead contact, Phd. Buwen Cao.

Materials availability
All materials reported in this paper will be shared by the lead contact upon request.

Data and code availability
Data reported in this paper will be shared by the lead contact upon request.
This paper does not report original code.
Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

METHOD DETAILS

MiRNA function similarity
The miRNA function similarity network was constructed based on a basic assumption that function-similar miRNAs tend to connect with similar diseases, and vice versa (Chen et al., 2018b; Goh et al., 2007; Lu et al., 2008). According to the valuable work of Wang et al. (Wang et al. (2010), miRNA function similarity data can be downloaded from http://www.cuilab.cn/files/images/cuilab/misim.zip. A miRNA function similarity network, including 484 miRNAs with 24061 interactions, was constructed using these data. It was derived from 5100 distinct experimentally confirmed associations between 326 diseases and 491 miRNAs after eliminating duplicate records. A total of 326 diseases containing associated miRNAs were used as the benchmark dataset to assess the results of the identified miRNA function modules.

MiRNA network similarity
A miRNA function similarity network is a complex network (Cao et al., 2021, 2022; Xiao et al., 2018). The present work analyzed and calculated the miRNA network similarity, which is described as follows:

\[
NetSim(m_i, m_j) = \frac{|N_{m_i} \cap N_{m_j}|}{\min(N_{m_i}, N_{m_j})}
\]  
(Equation 1)
Where NetSim(mi, mj) is the network similarity of the miRNA pair (mi, mj), denotes the first-level neighbors of miRNA mi.

**MiRNA combination similarity**

Using several features (e.g., similarities, weights, functions) can effectively compensate for the shortcoming of a single feature in a biological network (Cao et al., 2021, 2022; Chen et al., 2019a; Ding et al., 2018; Wang et al., 2021a) and improve the quality of predicted results. In this study, miRNA function similarity and network topological similarity were combined. The mathematical expression is as follows:

$$\text{ComSim}(mi, mj) = \frac{2 \times \text{MiSim}(mi, mj) + \text{NetSim}(mi, mj)}{\text{MiSim}(mi, mj) + \text{NetSim}(mi, mj)}$$  \hspace{1cm} (Equation 2)

Where ComSim(mi, mj) denotes the similarity of the miRNA pair (mi, mj), which is the harmonic mean between miRNA function similarity and network topology similarity, and MiSim(mi, mj) is the miRNA function similarity. The miRNA function interaction network is constructed after calculating the combined miRNA similarities. The average similarity of which is 0.64.

**Experimental models**

*ComSim-MINE methods*

Prior to presenting the ComSim-MINE algorithm, several terminologies relative to the ComSim-MINE method should be described.

A miRNA function interaction network can be described by the undirected weighted graph, $G = (V, E, \text{Sim})$, where node $V$ represents the miRNAs in the miRNA function interaction network, edge $E$ denotes the interaction between two different miRNAs, and Sim is the similarity value (i.e., weight) between two different miRNAs. Therefore, the density of a weighted subgraph of $S$ is described as follows:

$$d_{\text{Sim}}(S) = \frac{\sum \text{Sim}}{|V| \times ((|V| - 1)/2)}$$  \hspace{1cm} (Equation 3)

Where Sim is the similarity value for each edge in the subgraph of $S$, $|V|$ denotes the number of nodes (i.e., miRNAs) within it. The similarity of a miRNA node $v$ is expressed as follows:

$$v_{\text{Sim}}(S) = k_{\text{max}} \times d_{\text{Sim}}(S)$$  \hspace{1cm} (Equation 4)

Where $k_{\text{max}}$ denotes the degree of the miRNA node $v$ in the weighted subgraph of $S$, which describes the set of miRNA node $v$ and its first-level neighbors.

A new extending rule, namely, extending coefficient (EC), was developed to refine edges with low similarities during the procedure of the ComSim-MINE algorithm to further improve the identification accuracy of associations between diseases and miRNAs. The extending coefficient of $v_i$ to a miRNA function module $L_j$ is defined as follows:

$$EC(v_i, L_j) = \frac{\sum_{v_j \in L_j} \text{Sim}(v_i, v_j)}{|E(v_i, v_j)|}$$  \hspace{1cm} (Equation 5)

Where the numerator is the sum of the similarities between a miRNA and all miRNAs in the function module $L_j$, and the denominator is the edge number between a miRNA and all miRNAs in the function module $L_j$. If $EC(v_i, L_j) \geq \alpha$, which is predefined threshold, then $v_i$ is added to $L_j$. In our study, the value of $\alpha$ ranges from 0 to 0.9. The smaller the value of $\alpha$, the more neighbor miRNAs will be added to the miRNA function module. Generally, the optimal value of $\alpha$ can be determined by testing the value in the miRNA function interaction network. The below figure illustrates how to extend the miRNA function module based on EC.

Without a loss of generality, the similarity between a miRNA pair is computed based on Equation 1. Assume that $\alpha = 0.6$ and $L_1 = \{m_1, m_2, m_3, m_4\}$. For $m_5$, $EC(m_5, L_1) = (0.75 + 0.8)/2 = 0.725$, and for $m_6$,
EC(m6, L1) = 0.5/1 = 0.5. Therefore, miRNA m5 will be added to L1, and miRNA m6 will be excluded for L1. The final identified miRNA function module is \{m1, m2, m3, m4, m5\}.

**ComSim-MINE algorithm**

Similar to MINE [38], which is employed to generate protein complexes, the steps are as follows: first, MINE ranks proteins according to the vertex weight computed from edge degree and local neighborhood density; second, MINE generates protein complexes through iteratively maximizing the vertex weight or modularity score; finally, MINE merges the protein complexes that overlap by more than 50%, the ComSim-MINE algorithm consists of three steps. **Algorithm 1** presents the description of the ComSim-MINE algorithm.

First, ComSim-MINE calculates the similarity of each miRNA by **Equation 4** (lines 1 and 2) and then sorts the miRNAs in a descending manner (line 3).

**Algorithm 1. The Description of ComSim-MINE Algorithm**

Input: Graph \(G = (V, E)\).

\(vwp\): vertex similarity percentage;

\(msp\): modularity score percentage;

\(mp\): merge percentage;

\(\alpha\): the threshold of extending coefficient (EC)

Output: miRNA function module set \(L\)

1: Calculate the combined similarity of each edge in miRNA function interaction network by **Equation 2**;

2: Calculate the similarity of each miRNA in miRNA function interaction network by **Equation 4**, //S means the set of a node (\(v\)) and its first level neighbors;

3: Sort all miRNAs by the node similarities in the descending order;

4: \(L = \varnothing\);

5: for each seed miRNA, \(m_s \in M\), //Sorted by node similarity value in the decreasing order

6: \(L_t \in m_s\), //Add seed miRNA \(m_s\) to function module \(L_t\);

7: for \(N_m \in N_t\), //N: first level neighbor miRNA set of \(m_s\)

8: if \(EC \geq \alpha\) then

9: if \((v_{\infty}(S) of N_m) \geq (v_{\infty}(S) of m_s)(1-vwp)\) or modularity score \((L_t U N_m)\) > modularity-score \((L_t) + \) modularity score \((L)\) *msp // S means the set of a node and its first level neighbors
Second, ComSim-MINE establishes iteratively a miRNA function module through each of the seed miRNA generated in the first step (line 4) and expands it by absorbing neighbors if either of two conditions is met:

(i) The neighbor miRNA similarity is above the node similarity percentage (wp) of the seed miRNA, and the miRNA function module score is equal to or greater than the modularity score percentage (msp); and

(ii) The modularity score for the miRNA function module is improved by msp. Modularity score (Ms) is defined as follows:

$$Ms = \frac{E_{in}}{E_{out}}$$  \hspace{1cm} (Equation 6)

Where $E_{in}$ denotes the number of edges between miRNAs in a function module, and $E_{out}$ is the number of edges between module members and non-members. In this study, the EC defined in Equation 5 was employed to refine edges with low similarities in two conditions. If $EC \geq \alpha$, then $vi$ is added to $L_j$. Therefore, the candidate miRNA function will be produced by EC and either of the two conditions (lines 5–13). Finally, the candidate miRNA function modules are assessed for modularity score if miRNAs are deleted (lines 14–19).

Finally, ComSim-MINE merges candidate miRNA function modules with overlapping score $> 0.5$ (line 18–21) and removes the miRNA function modules with a size of less than 3 to guarantee the effectiveness of the experimental results (line 23).

Notably, different from the traditional MINE algorithm, we improved the original method from the following aspects: the first aspect is extending coefficient (EC), which is developed to refine edges with low similarities during the procedure of the ComSim-MINE algorithm to further improve the identification accuracy of associations between diseases and miRNAs; the second aspect is miRNA combination similarity, which calculated the harmonic mean of the miRNA function similarity and network topological similarity, effectively compensated for the shortcoming of a single feature in a biological network.