

MISIM v2.0: a web server for inferring microRNA functional similarity based on microRNA-disease associations

Jianwei Li^{1,2,*}, Shan Zhang^{1,†}, Yanping Wan¹, Yingshu Zhao¹, Jiangcheng Shi², Yuan Zhou² and Qinghua Cui^{2,3,*}

¹Institute of Computational Medicine, School of Artificial Intelligence, Hebei University of Technology, Tianjin 300401, China, ²Department of Biomedical Informatics, Department of Physiology and Pathophysiology, Center for Noncoding RNA Medicine, MOE Key Lab of Cardiovascular Sciences, School of Basic Medical Sciences, Peking University, 38 Xueyuan Rd, Beijing 100191, China and ³Sanbo Brain Institute, Sanbo Brain Hospital, Capital Medical University, Beijing 100093, China

Received January 27, 2019; Revised April 14, 2019; Editorial Decision April 22, 2019; Accepted April 25, 2019

ABSTRACT

MicroRNAs (miRNAs) are one class of important small non-coding RNA molecules and play critical roles in health and disease. Therefore, it is important and necessary to evaluate the functional relationship of miRNAs and then predict novel miRNA-disease associations. For this purpose, here we developed the updated web server MISIM (miRNA similarity) v2.0. Besides a 3-fold increase in data content compared with MISIM v1.0, MISIM v2.0 improved the original MISIM algorithm by implementing both positive and negative miRNA-disease associations. That is, the MISIM v2.0 scores could be positive or negative, whereas MISIM v1.0 only produced positive scores. Moreover, MISIM v2.0 achieved an algorithm for novel miRNA-disease prediction based on MISIM v2.0 scores. Finally, MISIM v2.0 provided network visualization and functional enrichment analysis for functionally paired miRNAs. The MISIM v2.0 web server is freely accessible at <http://www.liirmed.com/misim/>.

INTRODUCTION

MicroRNAs (miRNAs) are one class of important small non-coding RNA molecules which function as negative gene regulators by targeting mRNAs through base pairing (1,2). Given their important roles in various critical biological processes, it is well known that miRNAs are involved in a variety of human diseases, including cancer, cardiovascular diseases etc.(3), and thus could represent a novel therapeutic strategy (4–6). It was reported that miRNAs can be

clustered as functional sets (7,8), suggesting that miRNAs could be functionally related with other miRNAs. Thus, it is important to quantify the functional relations among miRNAs and then build miRNA functional networks.

For this purpose, a number of computational methods have been developed (9–14), for example miRFunSim (11) evaluates miRNA functional similarity based on the topological properties of protein-protein interaction (PPI) network (15) associated with the paired miRNAs, the method proposed by Yu *et al.* and MiRGOFS evaluate miRNA functional similarity based on GO semantic similarity metric (16) associated with the paired miRNAs. The MISIM algorithm, developed by us in 2010, is based on the increasingly accumulated miRNA-disease association data (12). Functional similarity is defined as a relationship, for example co-expression similarity, co-GO similarity, co-literature similarity, and co-similar disease similarity. MISIM first calculates disease similarity based on MeSH disease terms and then quantifies miRNA functional similarity by integrating miRNA-disease association data and disease similarity scores. Now MISIM has become a frequently used platform for addressing issues of miRNA-disease association prediction (17) and drug development (18). A number of recent studies have addressed the prediction of miRNA-disease association based on the original MISIM algorithm. Researchers presented some computational models for miRNA-disease association prediction, such as the models of Laplacian Regularized Sparse Subspace Learning (19), Extreme Gradient Boosting Machine (20), Bipartite Network Projection (21), Matrix Decomposition and Heterogeneous Graph Inference (22). As eight years have passed, however, limitations emerged for the original MISIM method and other algorithms. Besides MISIM v1.0

*To whom the correspondence should be addressed. Tel: +86 10 82801001; Fax: +86 10 82801001; Email: cuiqinghua@bjmu.edu.cn
Correspondence may also be addressed to Jianwei Li. Tel: +86 22 60435864; Fax: +86 22 60435867; Email: lijianwei@hebut.edu.cn

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

was not updated using the latest miRNA-disease association data, one major limitation is that MISIM v1.0 only produces scores ≥ 0 . All other algorithms also can only produce scores ≥ 0 . However, it is well known that some miRNAs are negatively related in function. For example, some miRNA activate cell apoptosis, but some others inhibit cell apoptosis. Therefore, the functional similarity score could be negative but MISIM v1.0 and all other algorithms failed to do so. In addition, MISIM v1.0 only produces miRNA functional similarity scores but not provides options such as network visualization, functional analysis, and novel miRNA-disease association prediction. Further tasks should be performed to implement these options. For example, users can further use software of Pajek or Cytoscape to visualize the produced miRNA functional networks. For functional analysis, they may use software of TAM (7,23) or miEAA (24). An integration of these options will greatly facilitate users.

To overcome the above limitations, we developed the MISIM v2.0 web server. By improving the algorithm of MISIM v1.0, MISIM v2.0 can quantify miRNA functional similarity scores to positive, negative, and zero values. In addition, MISIM v2.0 runs on the latest miRNA-disease association dataset from HMDD v3.0 (3) to keep a pace with the rate of data accrual. Moreover, MISIM v2.0 implemented network visualization and functional analysis for selected miRNAs. Finally, MISIM v2.0 implemented an algorithm to predict novel miRNA-disease association. Different with previous miRNA-disease association prediction algorithm, MISIM v2.0 can predict directional miRNA-disease associations, that is, it can determine the association between a miRNA and a disease is positive or negative.

METHODS

Collection of datasets

We downloaded MeSH terms from the National Library of Medicine (<http://www.nlm.nih.gov/>). The human miRNA-disease association dataset was obtained from the HMDD v3.0 database (<http://www.cuilab.cn/hmdd>) (3). The disease names in HMDD were curated using MeSH disease terms. Finally, we obtained 14250 miRNA-disease associations, including 1044 miRNAs and 613 diseases. The miRNA-disease association data was curated using the latest version of miRBase (Release 22.1). We obtained 547 miRNA sets from TAM 2.0 (<http://www.scse.hebut.edu.cn/tam/>) (7). Furthermore, all the source code of MISIM v2.0 was freely available.

Calculating miRNA functional similarity

The flowchart of MISIM v2.0 algorithm is shown in Figure 1. Firstly, we calculate the semantic value of a disease based on the MeSH disease structure of directed acyclic graph (DAG) with the method identical to the previous version of MISIM v1.0, $DAG_d = (d, T_d, E_d)$, Where d , T_d , E_d represent disease d , the set of all ancestor nodes of d including d itself and the set of corresponding links, respectively. The semantic value $D_d(t)$ of disease d can be calculated based

on the DAG_d .

$$\begin{cases} D_d(d) = 1 \\ D_d(t) = \max(\Delta * D_d(t') | t' \in \text{children of } t) \text{ if } t \neq d \end{cases} \quad (1)$$

where Δ is the semantic contribution factor for edges E_d linking disease t with its child disease t' . Based on equation (1), we calculate the semantic value of disease d as follows:

$$DV(d) = \sum_{t \in T_d} D_d(t) \quad (2)$$

The semantic similarity of two diseases is calculated in turn based on both the addresses of these diseases in DAG graphs and their semantic relations.

$$DS(d_1, d_2) = \frac{\sum_{t \in T_{d_1} \cap T_{d_2}} (D_{d_1}(t) + D_{d_2}(t))}{DV(d_1) + DV(d_2)} \quad (3)$$

After the above steps, the semantic similarity matrix of diseases is gained in the end and it is convenient to fetch the semantic similarity between arbitrary two diseases.

Secondly, it is known that miRNAs associated with similar diseases may have similar functions. Based on the miRNA-disease association from the datasets in MISIM v2.0, the matrix of semantic feature of the miRNA-disease association data is calculated. Moreover, we have in-depth curation of miRNA-disease associations by discriminating the up-regulated (positive) miRNAs and down-regulated (negative) miRNAs.

To better describe the method representing the up-regulated and down-regulated relations between miRNAs and diseases, we improve the formula (3), the semantic value of disease d is defined as

$$r_DV(d) = (-1)^r \sum_{t \in T_d} D_d(t) \quad (4)$$

where r represents the original deregulation types (up or down regulation), which has two possible values, +1 for down-regulated miRNAs and 0 for up-regulated miRNAs. Based on the above formulas, the miRNA-disease associations can be quantified as semantic feature vectors, which is defined as

$$f_{mi} = \{r_DV(d_{i1}), r_DV(d_{i2}), \dots, r_DV(d_{in})\} \quad (5)$$

where n is the total number of diseases associated with miRNA mi . Formula (5) reflects the global semantic feature of the diseases regulated by one miRNA in the MeSH disease structure of DAG. A bigger absolute value of one component in the disease vectors indicates the corresponding disease of this component includes more semantic information. It also means that the corresponding disease is more specific than other diseases which are regulated by the same miRNA.

At last, when the up-regulated and down-regulated miRNA-disease associations are considered, the difference of directions between two semantic feature vectors of miRNAs $m1$ and $m2$ should be taken into account when computing miRNA functional similarity. Then the association

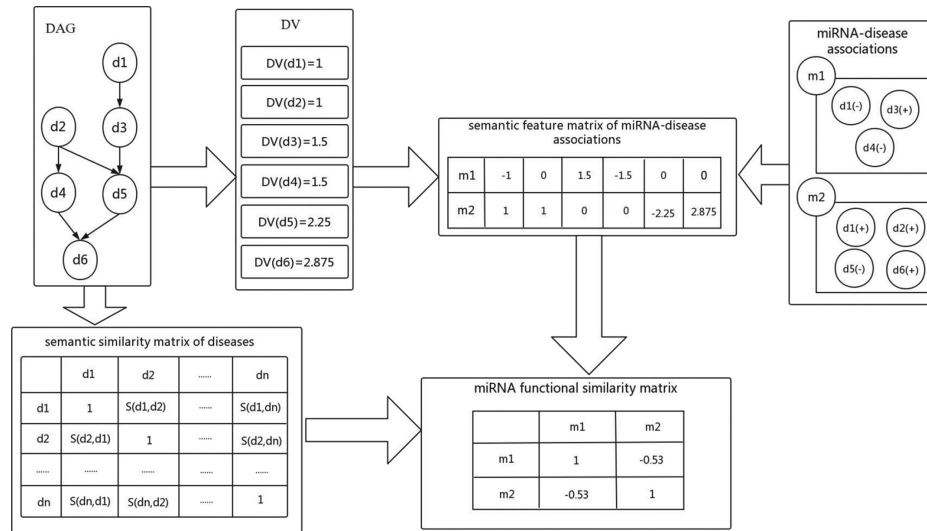


Figure 1. The flowchart of MISIM v2.0 algorithm.

score vectors of the input miRNAs $m1$ and $m2$ can be calculated by using cosine correlation as:

$$COS_{mir}(m1, m2) = \frac{f_{m1} f_{m2}^T}{\|f_{m1}\| \|f_{m2}\|} \quad (6)$$

When calculating miRNA functional similarity, we found a common phenomenon. The value of cosine correlation is often zero when the intersection of two disease sets of $m1$ and $m2$ is empty. However, we observed that there exist associations among MeSH disease structures of $m1$ and $m2$. This means that the value of cosine correlation should not be zero in this case.

Based on the above observation, formula (3) is adopted to calculate the semantic similarity of any two diseases associated with $m1$ and $m2$. Next, the disease sets associated with $m1$ and $m2$ are named as $D1$ and $D2$. Then, the disease semantic similarity of $m1$ and $m2$ is calculated by formula (7).

$$f_{MDS(m1, m2)} = \{DS(d_{11}, d_{21}), DS(d_{12}, d_{21}), \dots, DS(d_{1n}, d_{2m})\} \quad (7)$$

Then, the disease semantic features of $m1$ and $m2$ are improved by the following formulas respectively:

$$f_{m1} = \{f_{m1}, f_{MDS(m1, m2)}\} \quad (8)$$

$$f_{m2} = \{f_{m2}, f_{MDS(m2, m1)}\} \quad (9)$$

Finally, miRNA functional similarity will be calculated using formula (6) based on the novel disease semantic features calculated using formulas (8) and (9).

Predicting novel miRNA-disease associations

Moreover, MISIM v2.0 can also be used to predict novel miRNA-disease associations and to infer novel potential functions or associated diseases for given miRNAs. It provides two predicting methods, one is ‘miRNA-Disease Association’, and the other one is ‘miRNA-Disease Association for all’.

The method of ‘miRNA-Disease Association’ is to predict the novel miRNA-disease associations of two miRNAs, $m1$ and $m2$. The associated diseases sets of $m1$ and $m2$ are named $D1$ and $D2$, respectively. Then the intersection of $D1$ and $D2$ is calculated and is named as $D3$. Next, MISIM v2.0 calculates the subtraction between $D1$ and $D3$ and then the subtraction will be considered to be the predicted novel disease associated with $m2$. Moreover, the reciprocals of the disease semantic values are treated as the probability of potential novel diseases. A higher probability means the prediction result is more reliable. Normally, the general disease has a higher reciprocal of the semantic value.

Correspondingly, the method of ‘miRNA-Disease Association for all’ is to predict the novel miRNA-disease associations in the total datasets for single miRNA submitted by the user. In the process of prediction, the miRNA $m1$ submitted by the user is calculated with the miRNA mi ($mi \neq m1$) in the dataset one by one. The associated disease sets of $m1$ and mi are named $D1$ and Di , the intersection of $D1$ and Di was $D3$. The subtraction of Di and $D3$ will be the predicted novel miRNA-disease associations of $m1$. MISIM v2.0 takes the values of miRNA functional similarity as weighting factors of this subtraction, and calculates the weighted frequencies of each novel disease.

miRNA functional enrichment analysis

For calculated functionally paired miRNAs or miRNAs in a constructed miRNA functional network, it is interesting to dissect the enriched functions of these miRNAs. For doing so, we implemented the algorithm of TAM 2.0 (7) in MISIM v2.0.

Visualization of miRNA functional networks

We implemented the visualization of miRNA functional networks by visjs + G2. Users can change node shape (e.g. box, diamond, triangle, ellipse), node size, and set the threshold of functional similarity value for whether there is

a link between two miRNAs. In addition, if the users choose an option for directional functional similarity, the positive links (functional similarity with positive value) and negative links (functional similarity with negative value) will be in different colors. Once the network was finalized, users also can download the network picture.

Server construction

We built the MISIM v2.0 web server in the ‘Linux + Apache + MySQL + PHP’ framework. The data visualization was implemented by using open source visjs package (http://visjs.org/network_examples.html) and G2 package (<http://npm.taobao.org/package/g2>).

RESULTS

Overview of MISIM v2.0 web server

Currently, MISIM v2.0 contains 1044 miRNAs and 613 diseases, a 3-fold increase compared to MISIM v1.0. More importantly, it collects information that how deregulation (up-regulated or down-regulated) of a miRNA is involved in a corresponding disease.

MISIM v2.0 works as the following flowchart. Firstly, for analysis, users have four options, ‘ALL vs ALL Similarity’ (which supports to calculate the paired functional similarity values for a group of miRNAs), ‘One vs ALL Similarity’ (which supports to calculate the paired functional similarity values for a single miRNA with all other miRNAs), ‘miRNA-Disease Association’ (which predicts the associated disease for a paired miRNAs by evaluating the functional relations of the two miRNAs), and ‘Prediction miRNA disease association for all’ (which predicts the associated disease for a single miRNA by evaluating the functional relations of the given miRNA with all other miRNAs). In addition, the calculated miRNA functional similarity values can be downloaded in the ‘Download’ menu. For the option of ‘ALL vs ALL Similarity’, the users need to first input candidate miRNA list (one miRNA in each line). Then they can check or uncheck ‘Considering up/down-regulation in the similarity calculation’ box. If they leave this box unchecked, MISIM v2.0 will not consider the direction of miRNA deregulation. A ‘Considering up/down-regulation in the similarity calculation’ option means MISIM v2.0 will consider up- or down- regulation of miRNAs in associated diseases. Finally, when ‘Submit’ was clicked, the calculated functional similarity scores of paired miRNAs of the inputted miRNAs will be shown. For the calculated results, the users can further perform network visualization and functional enrichment analysis. Moreover, MISIM v2.0 can be also used to predict the potential novel miRNA-Disease associations of input miRNAs.

Example for analyzing functional similarity for a list of miRNAs

To show this function of MISIM v2.0 web server, we click the menu of ‘ALL versus ALL Similarity’. Then we used the sample miRNA list. Next we select the ‘Considering up/down-regulation in the similarity calculation’ option in step 2. After clicking ‘Submit’ button, the result was thus

shown in the right panel of the page. We noted that there are positive and negative functional similarity values for paired miRNAs. For example, the functional similarity value between hsa-let-7f and hsa-mir-107 is 0.48628593, whereas that between hsa-mir-107 and hsa-mir-103 is -0.19892149 . The users can further download the calculated results, visualize these miRNAs and functional similarity values as a network, and perform functional enrichment analysis. The calculated miRNA functional network is shown as Figure 2A. If the users select the default option, that is, no consideration for the up-/down-regulation of miRNAs in disease, the network will be as Figure 2B.

Example for predicting associated disease for a miRNA

To show this function of MISIM v2.0 web server, we click the menu of ‘Prediction miRNA disease association for all’. We use hsa-mir-29 as an example and select ‘Considering up/down-regulation in the similarity calculation’ in the second step. After clicking ‘Predict’, the predicted associated diseases were shown in the down panel of this page. As a result, 570 novel diseases associated with hsa-mir-29 were predicted. For each predicted disease, the significance and possible deregulation direction were also given. For example, MISIM v2.0 predicted that hsa-mir-29 is negatively associated with Mental Disorders (Weighted frequency = -0.26). Although we did not find any experimental confirmations for this prediction in PubMed but found a literature, which reported that hsa-mir-29 was significantly decreased in the serum of patients with Alzheimer’s disease. Given that Mental Disorders and Alzheimer’s disease are highly related (25), it could be believable that hsa-mir-29 is also generally negatively involved in Mental Disorders. In addition, we can download all the predicted results.

CONCLUSION

In conclusion, we developed MISIM v2.0, which significantly improved the original MISIM algorithm. For a list miRNAs, MISIM v2.0 will calculate the functional similarity for all paired miRNAs. For one single miRNA, MISIM v2.0 will calculate the functional similarity of the given miRNAs with all other miRNAs one by one. For both the above options, MISIM v2.0 can produce the miRNA functional networks and perform miRNA functional enrichment analysis for the investigated miRNAs. In addition, MISIM v2.0 also can predict novel miRNA-disease associations which could be non-directional or directional.

Of course, although MISIM v2.0 was improved significantly compared with MISIM v1.0, limitations still exist. One limitation is that the miRNA-disease association data is incomplete, which could produce bias when calculating miRNA functional similarity. Given that a number of computational methods for the prediction of novel miRNA-disease association are based on MISIM scores. Thus, there should be bias for the prediction of novel miRNA-disease association. Therefore, integrating other methods, for example those based on GO terms or PPI network topologies, could improve the calculation of miRNA functional similarity. Another limitation is that the network visualization of MISIM v2.0 web server could do not work well for

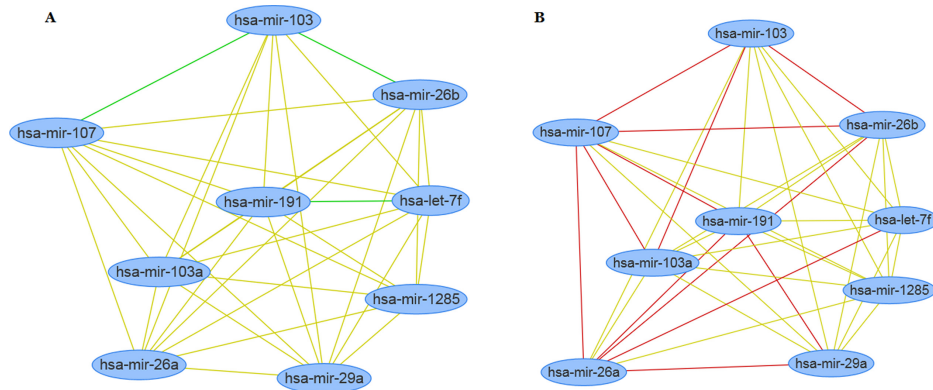


Figure 2. The visualization of the constructed miRNA functional networks using MISIM v2.0 web server and example miRNA list. (A) miRNA functional network with negative links produced using the option ‘Considering up/down-regulation in the similarity calculation’. (B) miRNA functional network without negative links produced using the option ‘default’.

some internet browser. The third limitation is that currently MISIM v2.0 only runs for pre-miRNAs and cannot discriminate between specific 5p and 3p mature miRNAs. The reason is that HMDD annotated miRNA-disease data at the level of miRNA precursor. We could improve HMDD and thus MISIM v2.0 by considering mature miRNAs as well. The fourth limitation is that currently MISIM v2.0 only runs for human miRNAs but cannot be applied to miRNAs from other species. The reason is also the HMDD database only contain human data. We will collect miRNA-disease association data from other species, for example, mouse and rat, and then update MISIM as well in the future. We will continue testing MISIM v2.0 in different internet browsers to make it more convenient to users.

ACKNOWLEDGEMENTS

We thank members at our groups for valuable discussions.

FUNDING

Special Project on Precision Medicine under the National Key R&D Program [2016YFC0903000 to Q.C.]; National Natural Science Foundation of China [81670462 to Q.C., 81672113 to J.L., 31801099 to Y.Z.]; Natural Science Foundation of Hebei Province [C2018202083 to J.L.]; Fundamental Research Funds for Central Universities of China [BMU2017YJ004 to Y.Z.]. Funding for open access charge: National Natural Science Foundation of China [81672113].

Conflict of interest statement. None declared.

REFERENCES

- Ambros, V. (2004) The functions of animal microRNAs. *Nature*, **431**, 350–355.
- Bartel, D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281–297.
- Huang, Z., Shi, J., Gao, Y., Cui, C., Zhang, S., Li, J., Zhou, Y. and Cui, Q. (2019) HMDD v3.0: a database for experimentally supported human microRNA-disease associations. *Nucleic Acids Res.*, **47**, D1013–D1017.
- Matsui, M. and Corey, D.R. (2017) Non-coding RNAs as drug targets. *Nat. Rev. Drug. Discov.*, **16**, 167–179.
- Rupaimoole, R. and Slack, F.J. (2017) MicroRNA therapeutics: towards a new era for the management of cancer and other diseases. *Nat. Rev. Drug. Discov.*, **16**, 203–222.
- Warner, K.D., Hajdin, C.E. and Weeks, K.M. (2018) Principles for targeting RNA with drug-like small molecules. *Nat. Rev. Drug. Discov.*, **17**, 547–558.
- Li, J., Han, X., Wan, Y., Zhang, S., Zhao, Y., Fan, R., Cui, Q. and Zhou, Y. (2018) TAM 2.0: tool for microRNA set analysis. *Nucleic Acids Res.*, **46**, W180–W185.
- Lu, M., Zhang, Q., Deng, M., Miao, J., Guo, Y., Gao, W. and Cui, Q. (2008) An analysis of human microRNA and disease associations. *PLoS One*, **3**, e3420.
- Ding, P., Luo, J., Xiao, Q. and Chen, X. (2016) A path-based measurement for human miRNA functional similarities using miRNA-disease associations. *Sci. Rep.*, **6**, 32533.
- Liu, Y., Zeng, X., He, Z. and Zou, Q. (2017) Inferring microRNA-disease associations by random walk on a heterogeneous network with multiple data sources. *IEEE/ACM Trans. Comput. Biol. Bioinf.*, **14**, 905–915.
- Sun, J., Zhou, M., Yang, H., Deng, J., Wang, L. and Wang, Q. (2013) Inferring potential microRNA-microRNA associations based on targeting propensity and connectivity in the context of protein interaction network. *PLoS One*, **8**, e69719.
- Wang, D., Wang, J., Lu, M., Song, F. and Cui, Q. (2010) Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases. *Bioinformatics*, **26**, 1644–1650.
- Yang, Y., Fu, X., Qu, W., Xiao, Y. and Shen, H.B. (2018) MiRGOFs: a GO-based functional similarity measurement for miRNAs, with applications to the prediction of miRNA subcellular localization and miRNA-disease association. *Bioinformatics*, **34**, 3547–3556.
- Yu, G., Xiao, C.L., Bo, X., Lu, C.H., Qin, Y., Zhan, S. and He, Q.Y. (2011) A new method for measuring functional similarity of microRNAs. *J. Integrated OMICS*, **1**, 49–54.
- Wang, Q., Sun, J., Zhou, M., Yang, H., Li, Y., Li, X., Lv, S., Li, X. and Li, Y. (2011) A novel network-based method for measuring the functional relationship between gene sets. *Bioinformatics*, **27**, 1521–1528.
- Xu, Y., Guo, M., Shi, W., Liu, X. and Wang, C. (2013) A novel insight into Gene Ontology semantic similarity. *Genomics*, **101**, 368–375.
- Chen, X., Wang, L., Qu, J., Guan, N.N. and Li, J.Q. (2018) Predicting miRNA-disease association based on inductive matrix completion. *Bioinformatics*, **34**, 4256–4265.
- Liang, X., Zhang, P., Yan, L., Fu, Y., Peng, F., Qu, L., Shao, M., Chen, Y. and Chen, Z. (2017) LRSSL: predict and interpret drug-disease associations based on data integration using sparse subspace learning. *Bioinformatics*, **33**, 1187–1196.
- Chen, X. and Huang, L. (2017) LRSSLMDA: Laplacian Regularized Sparse Subspace Learning for MiRNA-Disease Association prediction. *PLoS Comput. Biol.*, **13**, e1005912.

20. Chen,X., Huang,L., Xie,D. and Zhao,Q. (2018) EGBMMDA: Extreme Gradient Boosting Machine for MiRNA-Disease Association prediction. *Cell Death Dis.*, **9**, 3.
21. Chen,X., Xie,D., Wang,L., Zhao,Q., You,ZH. and Liu,H. (2018) BNPMDA: Bipartite Network Projection for MiRNA-Disease Association prediction. *Bioinformatics*, **34**, 3178–3186.
22. Chen,X., Yin,J., Qu,J. and Huang,L. (2018) MDHGI: Matrix Decomposition and Heterogeneous Graph Inference for miRNA–disease association prediction. *PLoS Comput. Biol.*, **14**, e1006418.
23. Lu,M., Shi,B., Wang,J., Cao,Q. and Cui,Q. (2010) TAM: a method for enrichment and depletion analysis of a microRNA category in a list of microRNAs. *BMC Bioinformatics*, **11**, 419.
24. Backes,C., Khaleeq,Q.T., Meese,E. and Keller,A. (2016) miEAA: microRNA enrichment analysis and annotation. *Nucleic Acids Res.*, **44**, W110–W116.
25. Ma,W., Wu,M., Zhou,S., Tao,Y., Xie,Z. and Zhong,Y. (2018) Reduced Smoothed level rescues Aβ-induced memory deficits and neuronal inflammation in animal models of Alzheimer’s disease. *J. Genet. Genomics*, **45**, 237–246.