

RESEARCH

Open Access



DisSetSim: an online system for calculating similarity between disease sets

Yang Hu¹, Lingling Zhao², Zhiyan Liu², Hong Ju³, Hongbo Shi⁴, Peigang Xu², Yadong Wang^{2*} and Liang Cheng^{4*}

From Biological Ontologies and Knowledge bases workshop on IEEE BIBM 2016
Shenzhen, China. 16 December 2016

Abstract

Background: Functional similarity between molecules results in similar phenotypes, such as diseases. Therefore, it is an effective way to reveal the function of molecules based on their induced diseases. However, the lack of a tool for obtaining the similarity score of pair-wise disease sets (SSDS) limits this type of application.

Results: Here, we introduce *DisSetSim*, an online system to solve this problem in this article. Five state-of-the-art methods involving Resnik's, Lin's, Wang's, PSB, and SemFunSim methods were implemented to measure the similarity score of pair-wise diseases (SSD) first. And then "pair-wise-best pairs-average" (PWBPA) method was implemented to calculate the SSDS by the SSD. The system was applied for calculating the functional similarity of miRNAs based on their induced disease sets. The results were further used to predict potential disease-miRNA relationships.

Conclusions: The high area under the receiver operating characteristic curve AUC (0.9296) based on leave-one-out cross validation shows that the PWBPA method achieves a high true positive rate and a low false positive rate. The system can be accessed from <http://www.bio-annotation.cn:8080/DisSetSim/>.

Keywords: Functional similarity, Similarity score, Disease sets, Disease-miRNA relationships

Background

The similarity of pair-wise disease sets (SDS) has drawn more and more attention in identifying functional similarity of the disease-caused molecules [1], predicting potential relationships between diseases and molecules [2–8], and so on. In previous studies, Wang et al. utilized the SDS to construct a human miRNA functional similarity network (MFSN) [1]. And Sun et al. used the SDS to predict novel disease lncRNA relationships [9].

The performance of calculating the SDS is mainly based on the method for computing the similarity of pair-wise diseases (SD). Currently, seven state-of-art methods involving Resnik's [10], Lin's [11], Wang's [12], process-similarity based (PSB) [13], SemFunSim [14], ILNCSIM [15], and FMLNCSIM [16] methods were

frequently used for computing the SD. Among these methods, Resnik's [10], Lin's [11], and Wang's methods [12] are designed earlier for Gene Ontology (GO) [8, 17]. And these methods were introduced for calculating the SD by DOSim [18] and DisSim [19]. Resnik's and Lin's methods [10, 11] are based on information content (IC) for computing similarity between terms of ontology. And Wang's method [12] is based on the hierarchical structure of the ontology. PSB and SemFunSim methods are newly developed for Disease Ontology (DO) [20]. PSB method [13] utilized the association of biological process between genes to calculate disease similarity. In comparison, SemFunSim method [14] considered more types of the functional associations including protein-protein interaction [21], human mRNA co-expression [22], and so on.

Resources for calculating the similarity score of pair-wise diseases (SSD) mainly includes the vocabularies of diseases and disease-related genes. The frequently used disease vocabularies contain Online Mendelian Inheritance in Man (OMIM) [23], Medical Subject Headings

* Correspondence: ydwang@hit.edu.cn; liangcheng@hrbmu.edu.cn
²Department of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, People's Republic of China
⁴College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150001, People's Republic of China
Full list of author information is available at the end of the article

(MeSH) [24], and DO [20]. OMIM records the names of genetic disorders without providing semantic associations between them. MeSH provides a hierarchy of terms in biomedical domain. It contains 16 categories, of which only C and F03 involve disease names. In comparison with OMIM and MeSH, DO has been established around the concept of disease, and it aims to provide a clear definition for each disease. The disease-related genes are scattered in the databases, such as Gene Reference into Function (GeneRIF) [25], OMIM [23], Genetic Association Database (GAD) [26] and Comparative Toxicogenomics Database (CTD) [27]. It is better to use relationships of all of these databases.

“pair-wise-all pairs-maximum” (PWAPM) method and “pair-wise-best pairs-average” (PWBPA) method are optional for calculating similarity of pair-wise term sets [28]. For comparing multiple aspects, the best measure is the PWBPA method, which is widely utilized in calculating similarity of DO and GO term sets [1, 7, 9, 12].

Although DOSim [18] and DisSim [19] implemented the disease similarity methods in R package and web interface, no tools provided the function to calculate the similarity score of pair-wise disease sets (SSDS) currently. In this article, we designed and implemented an online tool *DisSetSim* to calculate the SSDS. Five state-of-art disease similarity methods (Resnik’s, Lin’s, Wang’s, PSB, and SemFunSim) and the PWBPA method was implemented in the tool. The system is freely available at <http://www.bio-annotation.cn:8080/DisSetSim/>.

Methods

Date sources

Data sets of *DisSetSim* are from open source databases, and they are listed in Table 1. DO [20] records disease names. It provides terms for calculating disease similarity. GeneRIF [25], OMIM [23], GAD [26] and CTD [27] are manually curated databases of disease-related genes. All of diseases in these databases are mapped to terms in DO based on SIDD [29]. GO annotation (GOA) [30] includes functional annotation of genes. HumanNet is the gene functional network of human. In addition, HMDD v2.0 [31] contains disease-

Table 1 Data sources

Data source	Web site
DO	http://disease-ontology.org/
CTD	http://ctdbase.org/
GeneRIF	http://www.ncbi.nlm.nih.gov/gene/about-generif
GAD	https://geneticassociationdb.nih.gov/
OMIM	http://www.omim.org/
GO & GOA	http://www.geneontology.org
HumanNet	http://www.functionalnet.org/humannet/
OAHG	bio-annotation.cn/OAHG/

related miRNAs, diseases of which were manually mapped to terms in DO by OAHG [32].

Methods for calculating similarity score of pair-wise diseases

Five state-of-art methods involving Resnik’s [10], Lin’s [11], Wang’s [12], PSB [13], and SemFunSim methods [14] have been implemented for calculating the SSD.

Resnik’s and Lin’s methods are based on IC. The IC of a disease t is described as Eq. 1:

$$IC(t) = -\log_2 \frac{n_t}{N}, \tag{1}$$

where N is the total number of genes annotated by diseases, and n_t is the number of genes annotated by t . Assuming t_1 and t_2 are two diseases, the similarity of them is defined by Resnik as following [10]:

$$Sim_{Resnik}(t_1, t_2) = IC(t_{MICA}), \tag{2}$$

where t_{MICA} is the most informative common ancestor (MICA) of t_1 and t_2 . Lin defines the similarity of t_1 and t_2 as Eq. 3 [11]:

$$Sim_{Lin}(t_1, t_2) = \frac{2 \cdot IC(t_{MICA})}{IC(t_1) + IC(t_2)}. \tag{3}$$

Assuming T_1 is the set involving t_1 and all of its ancestor terms of ontology. Semantic contribution of term t to t_1 is represented as following:

$$S_{t_1}(t) = \begin{cases} 1 & t = t_1 \\ S_{t_1}(t) = \max\{w \cdot S_{t_1}(t') \mid t' \in T_1\} & t \neq t_1 \end{cases}, \tag{4}$$

where w is the contribution factor of each semantic relationship. According to Wang et al. [1], w is defined as 0.5 for ‘IS_A’ relationship of DO [20]. Then, all the semantic contributions of T_1 to t_1 is $SV(t_1)$, which is defined as following:

$$SV(t_1) = \sum_{t \in T_1} S_{t_1}(t). \tag{5}$$

Assuming T_2 is the set involving t_2 and all of its ancestor terms, the similarity between t_1 and t_2 is defined as following by Wang’s method [12]:

$$Sim_{Wang}(t_1, t_2) = \frac{\sum_{t \in T_1 \cap T_2} (S_{t_1}(t) + S_{t_2}(t))}{SV(t_1) + SV(t_2)}. \tag{6}$$

Assuming t_1 and t_2 can be related with m and n biological processes of GO based on hypergeometric test, respectively, the similarity of t_1 and t_2 is defined by the PSB method as following:

$$Sim_{PSB}(t_1, t_2) = \frac{1}{2} \left(\frac{\sum_{i=1}^m \max_{1 \leq j \leq n} (Sim(p_{1i}, p_{2j}))}{m} + \frac{\sum_{j=1}^n \max_{1 \leq i \leq m} (Sim(p_{2j}, p_{1i}))}{n} \right) \quad (7)$$

where p_{1i} and p_{2j} is the i th and j th significant related biological process terms of t_1 and t_2 , respectively. $Sim(p_{1i}, p_{2j})$ represents similarity between two processes p_{1i} and p_{2j} , which is defined as Eq. 8:

$$Sim(p_1, p_2) = \frac{1}{2} \cdot (IC_{GO}(p_1) + IC_{GO}(p_2)) \cdot \frac{n(p_1 \cap p_2)}{n(p_1 \cup p_2)} \cdot \frac{IC_{GO}(p_1)}{Max(IC_{GO})} \cdot \frac{IC_{DO}(p_1)}{Max(IC_{DO})} \cdot \frac{IC_{GO}(p_2)}{Max(IC_{GO})} \cdot \frac{IC_{DO}(p_2)}{Max(IC_{DO})}, \quad (8)$$

where IC_{GO} and IC_{DO} represent IC based on GO and DO, respectively. $n(p_1 \cap p_2)$ and $n(p_1 \cup p_2)$ denote the number of common genes of p_1 and p_2 , and the number of total genes of p_1 and p_2 , respectively.

Assuming G_1 and G_2 represent related gene sets of t_1 and t_2 , respectively. Then, the similarity of t_1 and t_2 by the SemFunSim method can be described as following:

$$Sim_{SemFunSim}(t_1, t_2) = \frac{\sum_{i=1}^m \max_{1 \leq j \leq n} (Sim(g_{1i}, g_{2j})) + \sum_{j=1}^n \max_{1 \leq i \leq m} (Sim(g_{2j}, g_{1i}))}{\frac{m}{|G_{MICA}|} \cdot \frac{m+n}{n}} \quad (9)$$

where $|G_{MICA}|$ represents the number of genes in G_{MICA} . m and n denote the number of genes in G_1 and G_2 , respectively. $Sim(g_{1i}, g_{2j})$ is the functional similarity score between genes g_{1i} and g_{2j} , which could be obtained from HumanNet [33].

Method for calculating similarity score of pair-wise disease sets

The PWBPA method was utilized for calculating the SSDS. The similarity of two disease sets T_1 and T_2 is defined as following:

$$PWBPA(T_1, T_2) = \frac{\sum_{i=0}^N \max_{0 < j \leq M} Sim(t_i, t_j) + \sum_{j=0}^M \max_{0 < i \leq N} Sim(t_j, t_i)}{N + M}, \quad (10)$$

where T_1 and T_2 contains N and M diseases,

respectively. t_i and t_j represents i th and j th terms of T_1 and T_2 , respectively.

Predicting potential association between diseases and miRNAs

Functional similarity between miRNAs could be calculated based on their related disease sets. Similarities of each pair-wise miRNAs are utilized to establish a MFSN. Node of the network represents miRNA. Weight of edge is the functional similarity score. Then, disease-related miRNAs were prioritized using the network ranking algorithm named random walk with restart (RWR) [7].

The random walker starts on one or several seed nodes and then randomly transits to neighboring nodes considering the probabilities of the edges between the two nodes. And the probability to return to the seed nodes is supposed as γ . Then, RWR algorithm can be defined as following:

$$P_{t+1} = \gamma P_0 + (1-\gamma)AP_t \quad (11)$$

where P_0 denotes the initial probability vector, P_t is a vector in which the i th element represents the probability of finding the walker at node i and step t , A is the column-normalized adjacency matrix of the network. The algorithm was performed until the difference between P_t and P_{t+1} falling below 10^{-10} , which means all the nodes become stable.

In this study, the known miRNAs of a disease were considered as seed nodes. The unknown miRNAs of it could be scored based on RWR on the MFSN. After ranking the miRNAs based on the scores, disease-related miRNAs could be prioritized.

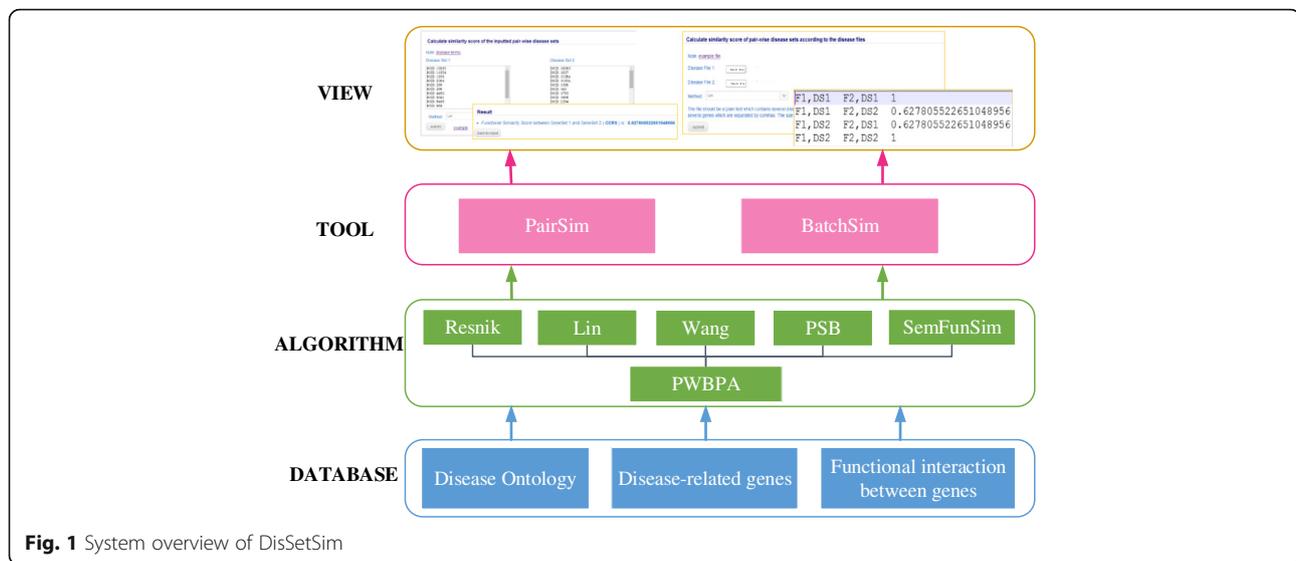
Implementation

DisSetSim has been implemented on a JavaEE framework and run on the web server (2-core (2.26 GHz) processors) of Ucloud [34]. The four-layer architecture involving DATABASE, ALGORITHM, TOOLS, and VIEW layer is shown in Fig. 1 The detailed description of the architecture is fixed as following.

(1) DATABASE layer. This layer stores DO, disease-related genes, and functional associations between genes. These are exploited by ALGORITHM layer for calculating the similarity between disease sets.

(2) ALGORITHM layer. Five algorithms of measuring the similarity between DO terms have been implemented, which include Resnik's, Lin's, Wang's, PSB, and SemFunSim methods. And the method named PWBPA for calculating the SSDS were also implemented.

(3) TOOL layer. Two tools including PairSim and BatchSim have been provided for exploring the SSDS. PairSim calculates the similarity for a given pair of



disease sets, and BatchSim computes similarity between each pair of multiple disease sets.

(4) VIEW layer. Web pages are provided for viewing the results. It shows the similarity of pair-wise disease sets.

Results

Web interface

DisSetSim provides two tools PairSim and BatchSim for querying the SSDS. The details about the usage of these two tools are described as follows.

The usage of PairSim

Figure 2a shows a case for searching the similarity score of a given pair of disease sets. The web page for inputting disease sets is <http://www.bio-annotation.cn:8080/DisSetSim/basic-init>. Each of these disease sets could be inputted in a textbox. A disease set is comprised by several diseases. And each disease is represented by the identifier of term in DO. All the term identifiers could be downloaded from the hyperlink 'disease terms' in the inputting page. Here, we click the 'example' button to use our example. Then, we choose one of the five methods (Resnik's, Lin's, Wang's, PSB, and SemFunSim) for calculating the SSD. After submitting this pair of disease sets, the system could return the similarity score based on the PWBPA method.

The usage of BatchSim

Figure 2b shows a case for searching the similarity score of all the pairs based on the selected files. The web page for inputting disease sets is <http://www.bio-annotation.cn:8080/DisSetSim/batch-init>. Two files including disease sets should be selected before submitting. The file should be a plain text which contains several disease

sets. Each disease set must be in a newline, and each disease set contains several disease IDs which are separated by commas. The size of uploaded file must be <2 Mb. Here, we selected our example file in this page. Then, we choose one of the five methods for calculating the SSD. After clicking the 'submit' button, the system could return the similarity score of all the pairs of the selected files based on the PWBPA method.

miRNA functional similarity network

By applying *DisSetSim* to the inputted disease sets of miRNAs, the similarity score of each miRNA pair could be obtained. Using miRNA as node and similar miRNAs as edge, the MFSN was constructed based on various similarity cutoffs. As shown in Fig. 3a, the number of links dramatically decreases when the cutoff increases from low value to high value. When the cutoff is equal to or bigger than 0.7, the link numbers remain relatively stable. Therefore, we use 0.7 as cutoff for the MFSN. In total, 1042 miRNA-miRNA functional associations between 346 miRNAs were obtained as MFSN (Fig. 3c). Similar to the most of the reported biological networks, the degree of this MFSN also shows a scale-free distribution [5, 9, 35–37]. It means that most of the miRNAs only have a few functionally similar miRNAs, and a few of miRNAs have a numerous functional similar miRNA (Fig. 3b).

Here, the PWBPA method was utilized for calculating similarity between disease sets, and SemFunSim was used as computing the similarity of pair-wise diseases. This is because that the SemFunSim method was proven to obtain the best performance [14]. Alternatively, other state-of-art methods could also be chosen to construct MFSN.

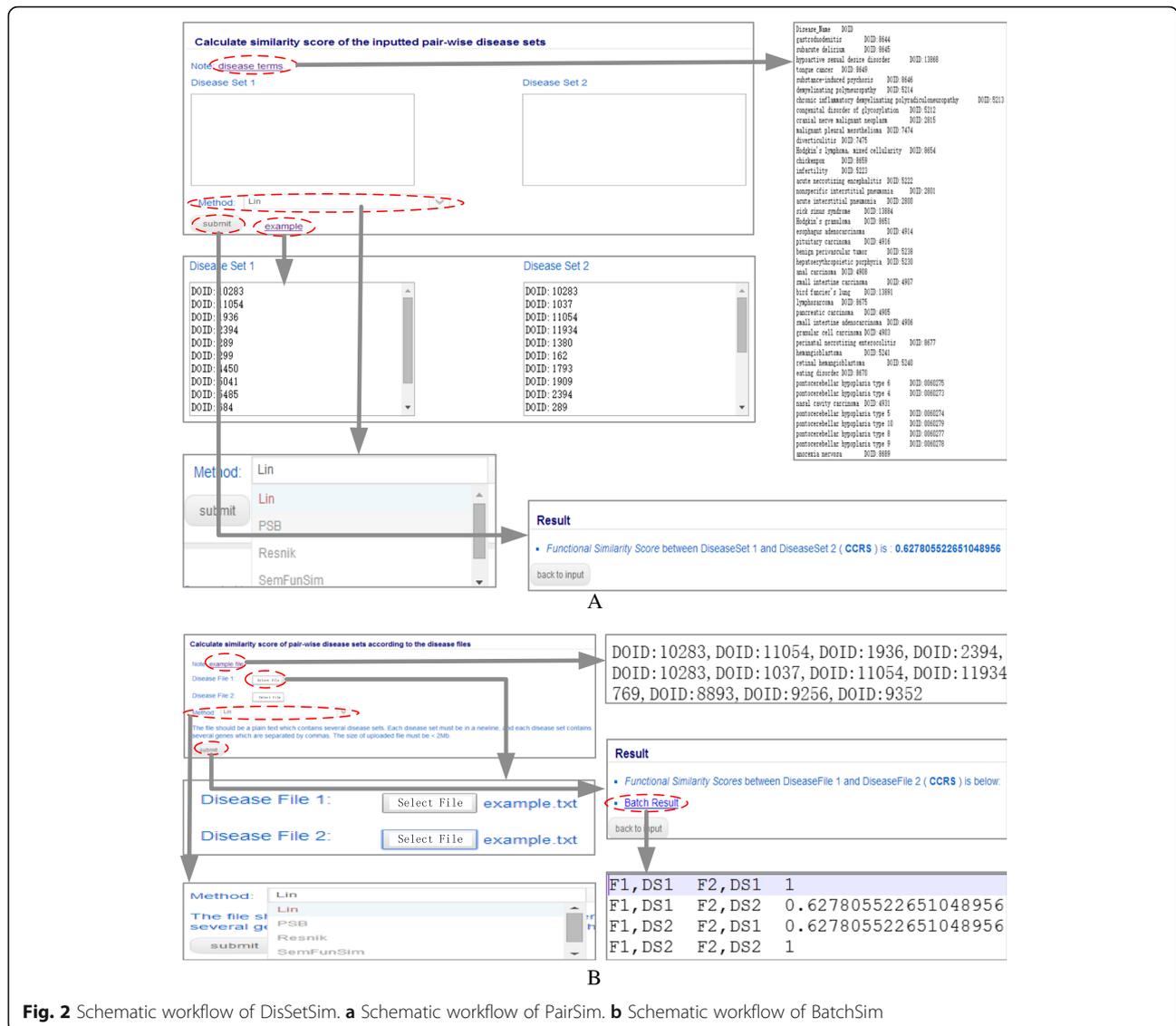


Fig. 2 Schematic workflow of DisSetSim. **a** Schematic workflow of PairSim. **b** Schematic workflow of BatchSim

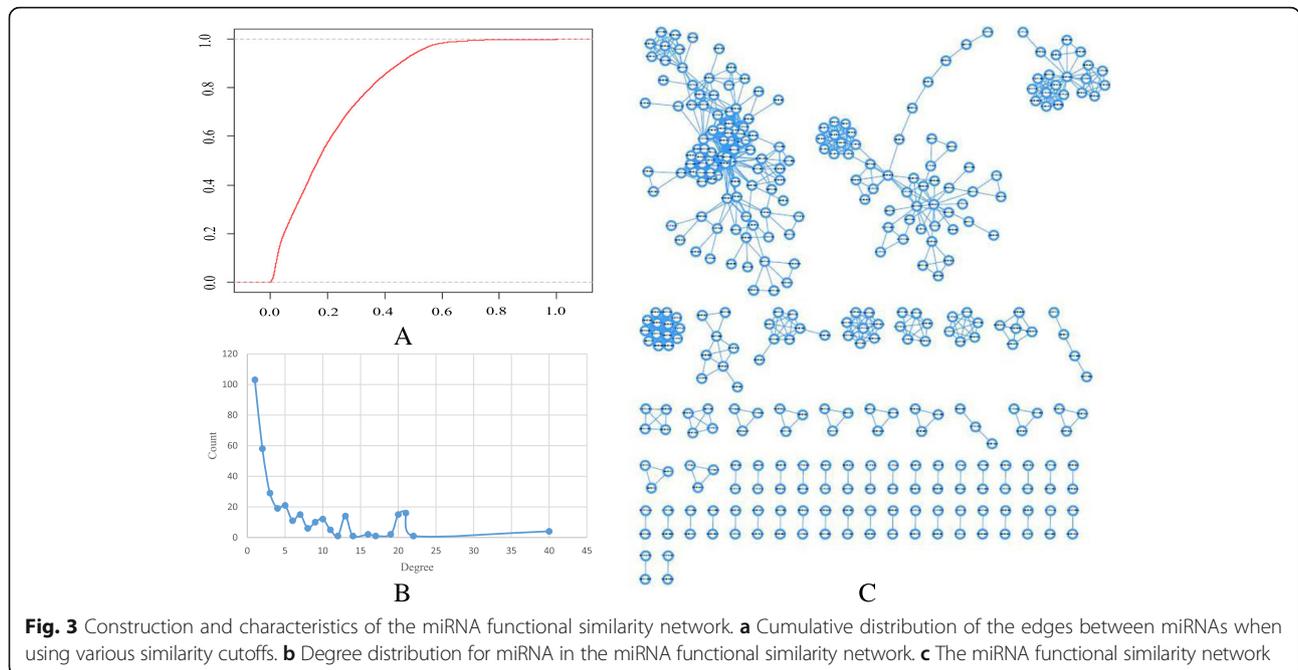
Disease-related miRNAs

By applying the above similarity scores of miRNAs, novel disease-related miRNAs were predicted based on RWR algorithm (See ‘Methods’ section). To evaluate the performance of the similarity scores of miRNAs, leave-one-out cross validation of 5710 known experimentally confirmed miRNA-disease associations, including 265 diseases with at least two miRNAs, were used for this assessment. For a disease of interest, each known miRNA of this disease was left out as the testing case, and the remaining miRNAs of this disease were used as seed nodes. All the miRNAs except the miRNAs of this disease were considered as candidate miRNAs. We then examined how well the testing miRNA ranked relative to the candidate miRNAs. If the ranking of this testing miRNA exceeded a

given cutoff, we regarded this miRNA-disease association as successfully predicted. As a result, an area under the ROC curve (AUC) of 0.9296 was achieved (Fig. 4), which demonstrated that our miRNA functional similarity was effective in recovering known experimentally confirmed disease-related miRNAs.

Discussion

As the best of our knowledge, non-coding RNAs (ncRNAs) attract more and more attentions because of their important regulation roles in molecular level. However, the lack of protein limits the identification of their function. Here the application of our tool in constructing MFSN and predicting miRNA-disease associations provides a novel way to help for exploring the function of miRNAs especially for prioritizing



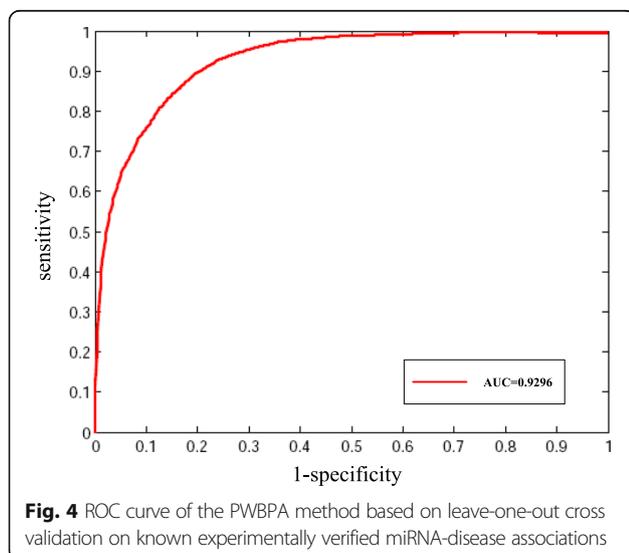
miRNA-disease associations. This application can be extended to other ncRNAs, such as lncRNAs and circRNAs. Although methods for calculating the SDS have been implemented by previous methods, it is not easy to calculate the SSDS. Therefore, DisSetSim benefits researchers for exploring the function of disease-related molecular.

Conclusions

In this article, we designed and developed a web system *DisSetSim* to calculate the SSDS. Five state-of-art methods were implemented (see ‘METHODS’ section)

for calculating disease similarity. And the PWBPA method was implemented for calculating the SSDS. Two tools involving PairSim and BatchSim provide the function to obtain the SSDS by inputting a pair-wise disease sets and multiple disease sets, respectively.

The functional similarity of miRNAs could be calculated based on our system. Here, the similarity of each pair-wise miRNAs was calculated. And then a MFSN was constructed based on miRNA similarity. The network was further utilized to predicate disease-related miRNAs based on RWR. The high AUC (0.9296) shows the MFSN is very suitable for predicting potential relationships between diseases and miRNAs.



Abbreviations

CTD: Comparative Toxicogenomics Database; DO: Disease Ontology; GAD: Genetic Association Database; GeneRIF: Gene Reference into Function; GO: Gene Ontology; GOA: GO annotation; IC: information content; MFSN: miRNA functional similarity network; MICA: the most informative common ancestor; ncRNA: non-coding RNAs; OMIM: Online Mendelian Inheritance in Man; PSB: process-similarity based; PWAPM: pair-wise-all pairs-maximum; PWBPA: pair-wise-best pairs-average; PWR: random walk with restart; SD: pair-wise diseases; SSD: the similarity score of pair-wise diseases; SSDS: the similarity score of pair-wise disease sets.

Acknowledgments

Yadong Wang and Liang Cheng are the corresponding authors. Lingling Zhao and Zhiyan Liu are the co-first authors.

Funding

This work was supported by the Major State Research Development Program of China [No. 2016YFC1202302], the National Natural Science Foundation of China (Grant No. 61502125, \$2000), Heilongjiang Postdoctoral Fund (Grant No. LBH-Z15179, \$800), and China Postdoctoral Science Foundation (Grant No. 2016 M590291, \$1000).

Availability of data and materials

The system can be accessed from <http://www.bio-annotation.cn:8080/DisSetSim/>.

About this supplement

This article has been published as part of *Journal of Biomedical Semantics* Volume 8 Supplement 1, 2017: Selected articles from the Biological Ontologies and Knowledge bases workshop. The full contents of the supplement are available online at <https://jbiomedsem.biomedcentral.com/articles/supplements/volume-8-supplement-1>.

Authors' contributions

YH, LZ and ZL implemented the first version of the online system. HJ, HS, PX updated the system. YW and LC wrote the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Harbin Institute of Technology, School of Life Science and Technology, Harbin 150001, People's Republic of China. ²Department of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, People's Republic of China. ³Department of information engineering, Heilongjiang Biological Science and Technology Career Academy, Harbin 150001, People's Republic of China. ⁴College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150001, People's Republic of China.

Published: 20 September 2017

References

- Wang D, Wang J, Lu M, Song F, Cui Q. Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases. *Bioinformatics*. 2010;26(13):1644–50.
- Chen X, Yan CC, Luo C, Ji W, Zhang Y, Dai Q. Constructing lncRNA functional similarity network based on lncRNA-disease associations and disease semantic similarity. *Sci Rep*. 2015;5:11338.
- Chen X, Yan CC, Zhang X, You Z-H. Long non-coding RNAs and complex diseases: from experimental results to computational models. *Brief Bioinform*. 2016;bbw060.
- Chen X, You Z, Yan G, Gong D. IRWRLDA: improved random walk with restart for lncRNA-disease association prediction. *Oncotarget*. 2016;7(36):57919–31.
- Peng J, Wang T, Hu J, Wang Y, Chen J. Constructing networks of organelle functional modules in Arabidopsis. *Current Genomics*. 2016;17(5):427–38.
- J. Peng, Bai, K., Shang, X., Wang, G., Xue, H., Jin, S., Cheng, L., Wang, Y., & Chen, J., "Predicting Disease-related Genes using Integrated Biomedical Networks," *BMC Genomics*, vol. 18, no. 1, pp. 1043, 2017.
- J. Peng, H. Li, Y. Liu, L. Juan, Q. Jiang, Y. Wang, and J. Chen, "InteGO2: a web tool for measuring and visualizing gene semantic similarities using Gene Ontology," *BMC Genomics*, vol. 17 Suppl 5, pp. 530, 2016.
- J. Peng, T. Wang, J. Wang, Y. Wang, and J. Chen, "Extending gene ontology with gene association networks," *Bioinformatics*, vol. 32, no. 8, pp. 1185–94, Apr 15, 2016.
- Sun J, Shi H, Wang Z, Zhang C, Liu L, Wang L, He W, Hao D, Liu S, Zhou M. Inferring novel lncRNA-disease associations based on a random walk model of a lncRNA functional similarity network. *Mol BioSyst*. 2014;10(8):2074–81.
- P. Resnik, "Using information content to evaluate semantic similarity in a taxonomy," *arXiv preprint cmp-lg/9511007*, 1995.
- Lin D. An information-theoretic definition of similarity. 1998. pp. 296–304.
- J. Z. Wang, Z. Du, R. Payattakool, P. S. Yu, and C. F. Chen, "A new method to measure the semantic similarity of GO terms," *Bioinformatics*, vol. 23, no. 10, pp. 1274–81, May 15, 2007.
- Mathur S, Dinakarbandian D. Finding disease similarity based on implicit semantic similarity. *J Biomed Inform*. 2012;45(2):363–71.
- L. Cheng, J. Li, P. Ju, J. Peng, and Y. Wang, "SemFunSim: a new method for measuring disease similarity by integrating semantic and gene functional association," *PLoS One*, vol. 9, no. 6, pp. e99415, 2014.
- Huang Y, Chen X, You Z, Huang D, Chan K. ILNCSIM: improved lncRNA functional similarity calculation model. *Oncotarget*. 2016;7(18):25902–14.
- Chen X, Huang Y, Wang X, You Z, Chan K. FMLNCSIM: fuzzy measure-based lncRNA functional similarity calculation model. *Oncotarget*. 2016;7(29):45948–58.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene ontology: tool for the unification of biology. The gene ontology consortium. *Nat Genet*. 2000;25(1):25–9.
- Li J, Gong B, Chen X, Liu T, Wu C, Zhang F, Li C, Li X, Rao S, Li X. DOSim: an R package for similarity between diseases based on disease ontology. *BMC Bioinformatics*. 2011;12:266.
- L. Cheng, Y. Jiang, Z. Wang, H. Shi, J. Sun, H. Yang, S. Zhang, Y. Hu, and M. Zhou, "DisSim: an online system for exploring significant similar diseases and exhibiting potential therapeutic drugs," *Sci Rep*, vol. 6, pp. 30024, Jul 26, 2016.
- Schriml LM, Arze C, Nadendla S, Chang YW, Mazaitis M, Felix V, Feng G, Kibbe WA. Disease ontology: a backbone for disease semantic integration. *Nucleic Acids Res*. 2012;40(Database issue):D940–6.
- Ortutay C, Vihinen M. Identification of candidate disease genes by integrating gene Ontologies and protein-interaction networks: case study of primary immunodeficiencies. *Nucleic Acids Res*. Feb, 2009;37(2):622–8.
- J. M. Stuart, E. Segal, D. Koller, and S. K. Kim, "A gene-coexpression network for global discovery of conserved genetic modules," *Science*, vol. 302, no. 5643, pp. 249–55, Oct 10, 2003.
- Amberger J, Bocchini C, Hamosh A. A new face and new challenges for online Mendelian inheritance in man (OMIM®). *Hum Mutat*. 2011;32(5):564–7.
- Dhammi IK, Kumar S. Medical subject headings (MeSH) terms. *Indian J Orthop*. 2014;48(5):443–4.
- Mitchell JA, Aronson AR, Mork JG, Folk LC, Humphrey SM, Ward JM. Gene indexing: characterization and analysis of NLM's GeneRIFs. *AMIA Annu Symp Proc*. 2003:460–4.
- Becker KG, Barnes KC, Bright TJ, Wang SA. The genetic association database. *Nat Genet*. May, 2004;36(5):431–2.
- Davis AP, Murphy CG, Johnson R, Lay JM, Lennon-Hopkins K, Saraceni-Richards C, Sciaky D, King BL, Rosenstein MC, Wiegers TC, Mattingly CJ. The comparative Toxicogenomics database: update 2013. *Nucleic Acids Res*. Jan, 2013;41(Database issue):D1104–14.
- C. Pesquita, D. Faria, A. O. Falcao, P. Lord, and F. M. Couto, "Semantic similarity in biomedical ontologies," *PLoS Comput Biol*, vol. 5, no. 7, pp. e1000443, 2009.
- L. Cheng, G. Wang, J. Li, T. Zhang, P. Xu, and Y. Wang, "SIDD: A Semantically Integrated Database towards a Global View of Human Disease," *PLOS ONE*, vol. 8, no. 10, pp. e75504, 2013.
- Camon E, Magrane M, Barrell D, Lee V, Dimmer E, Maslen J, Binns D, Harte N, Lopez R, Apweiler R. The gene ontology annotation (goa) database: sharing knowledge in uniprot with gene ontology. *Nucleic Acids Res*. 2004;32(suppl 1):D262–6.
- Li Y, Qiu C, Tu J, Geng B, Yang J, Jiang T, Cui Q. HMDD v2.0: a database for experimentally supported human microRNA and disease associations. *Nucleic Acids Res*. 2014;42(Database issue):D1070–4.
- Cheng L, Sun J, Xu W, Dong L, Hu Y, Zhou M. OAHG: an integrated resource for annotating human genes with multi-level ontologies. *Sci Rep*. 2016;10:34820.
- Lee I, Blom UM, Wang PI, Shim JE, Marcotte EM. Prioritizing candidate disease genes by network-based boosting of genome-wide association data. *Genome Res*. 2011;21(7):1109–21.
- Sqalli MH, Al-Saeedi M, Binbeshr F, Siddiqui M. UCloud: A simulated Hybrid Cloud for a university environment. 2012. pp. 170–172.
- K. I. Goh, M. E. Cusick, D. Valle, B. Childs, M. Vidal, and A. L. Barabasi, "The human disease network," *Proc Natl Acad Sci U S A*, vol. 104, no. 21, pp. 8685–90, May 22, 2007.
- Zhang F, Ren C, Lau KK, Zheng Z, Lu G, Yi Z, Zhao Y, Su F, Zhang S, Zhang B, Sobie EA, Zhang W, Walsh MJ. A network medicine approach to build a comprehensive atlas for the prognosis of human cancer. *Brief Bioinform*, Aug, 2016;24
- F. Zhang, B. Gao, L. Xu, C. Li, D. Hao, S. Zhang, M. Zhou, F. Su, X. Chen, H. Zhi, and X. Li, "Allele-specific behavior of molecular networks: understanding small-molecule drug response in yeast," *PLoS One*, vol. 8, no. 1, pp. e53581, 2013.