

RESEARCH ARTICLE

Open Access



Prediction and interpretation of miRNA-disease associations based on miRNA target genes using canonical correlation analysis

Hailin Chen^{1*} , Zuping Zhang² and Dayi Feng¹

Abstract

Background: It has been shown that the deregulation of miRNAs is associated with the development and progression of many human diseases. To reduce time and cost of biological experiments, a number of algorithms have been proposed for predicting miRNA-disease associations. However, the existing methods rarely investigated the cause-and-effect mechanism behind these associations, which hindered further biomedical follow-ups.

Results: In this study, we presented a CCA-based model in which the possible molecular causes of miRNA-disease associations were comprehensively revealed by extracting correlated sets of genes and diseases based on the co-occurrence of miRNAs in target gene profiles and disease profiles. Our method directly suggested the underlying genes involved, which could be used for experimental tests and confirmation. The inference of associated diseases of a new miRNA was made by taking into account the weight vectors of the extracted sets.

We extracted 60 pairs of correlated sets from 404 miRNAs with two profiles for 2796 target genes and 362 diseases. The extracted diseases could be considered as possible outcomes of miRNAs regulating the target genes which appeared in the same set, some of which were supported by independent source of information. Furthermore, we tested our method on the 404 miRNAs under the condition of 5-fold cross validations and received an AUC value of 0.84606. Finally, we extensively inferred miRNA-disease associations for 100 new miRNAs and some interesting prediction results were validated by established databases.

Conclusions: The encouraging results demonstrated that our method could provide a biologically relevant prediction and interpretation of associations between miRNAs and diseases, which were of great usefulness when guiding biological experiments for scientific research.

Keywords: miRNA-disease associations, Target genes, Canonical correlation analysis

Background

microRNAs (miRNAs) are one category of small non-coding RNAs that regulate gene expressions by base pairing with 3'-UTRs of messenger RNAs (mRNAs). Since the initial discovery in 1993 [1], the number of currently annotated miRNAs has increased steadily. As of November 2018, the newest version of miRBase [2] contained >480,000 mature miRNA sequences in 271

species. There are growing studies suggesting that miRNAs play important roles in some essential biological processes, such as cell proliferation [3], development [4], differentiation [5, 6] and metabolism [7]. Hence the dysfunction of miRNAs will result in aberrant cell behaviors, and they have been associated with the development and progression of many human diseases. For example, Nagaraja et al. [8] revealed that the overexpression of *mir-100* inhibited mTOR signaling in clear cell ovarian cancer. In addition, the conserved sequences, specific secondary structures and the ability to

* Correspondence: chl_mail@csu.edu.cn

¹School of Software, East China Jiaotong University, Nanchang 330013, China
Full list of author information is available at the end of the article



control gene expression makes miRNAs suitable targets for drug development [9] and recent studies [10–13] have demonstrated their application in the therapeutic exploitation.

Because of the wide-spread clinical implications, some online databases [14–16] have been established for containing experimentally confirmed evidence for associations between miRNAs and diseases via text mining. These repositories serve as comprehensive resources for studying the impacts of miRNAs on human diseases. However, our current knowledge of the involvement of miRNAs in diseases is far from completeness and thus those undiscovered associations cannot be mined from the literature. Meanwhile, experimental identification of miRNA-induced diseases by biological technology is costly and laborious. Therefore, computational prediction of the most promising miRNA-disease associations for further confirmation is receiving enormous attention.

The computational efforts made in this field can mainly be divided into two groups. The methods in the first group extensively exploited the biological evidence that miRNAs exert their functions by regulating the expression levels of their target mRNAs [17]. They first comprehensively collected two sets of genes, namely miRNA target genes and disease-related genes. The miRNA-disease associations were then inferred according to the similarity values or the interactions between the two sets of genes. For example, Mørk et al. [18] proposed a model miRPD to explicitly infer miRNA-protein-disease associations by coupling miRNA-gene associations with gene-disease associations. Li et al. [19] presented a computational framework to prioritize human cancer miRNAs by calculating the functional consistency scores (FCS) between the miRNA target genes and the cancer-related genes. Shi et al. [20] developed a computational method to identify potential miRNA-disease associations by mapping disease genes and miRNA target genes onto PPI (protein-protein interaction) networks for enrichment score calculation. These methods widely used predicted miRNA target genes to support miRNA-disease association inference. Due to high false positive rate of the predicted target genes of miRNAs, it is difficult for the above methods to achieve stable prediction results.

The methods in the other group were based on the conclusion that functionally similar miRNAs are usually involved in phenotypically similar diseases [21]. For instance, Chen et al. [22] developed a method RWRMDA to infer potential miRNA-disease associations by implementing random walks on miRNA-miRNA functional similarity network. Similarly, Chen et al. [23] devised a method to infer OMIM disease candidates related to a specific miRNA via random walks on disease similarity network. Afterwards, several methods [24–38] have been

presented to infer novel miRNA-disease associations by incorporating both the miRNA similarity network and the disease similarity network. Experimental results demonstrated that combining the two types of similarity networks could improve prediction performance. However, miRNA functional similarity calculation is a major challenge needed to be properly addressed in these methods. The functional similarity between two miRNAs was usually measured based on their associated diseases, which could produce overestimated validation accuracy [37]. Moreover, these similarity-based methods do not explicitly facilitate forming hypotheses about the possible molecular causes of the miRNA-disease associations [18].

In this paper, we presented a novel method to predict potentially related diseases of miRNA candidates based on their target genes on a large scale, without limiting ourselves to similarity measurement. Experimentally supported miRNA-gene interactions and miRNA-disease associations were first obtained from existing databases to build target gene profiles and disease profiles for miRNAs. We then applied canonical correlation analysis (CCA) to extract correlated sets of genes and diseases based on the co-occurrence of miRNAs in the two profiles. For a new miRNA, its disease profiles were inferred based on the weight vectors of the extracted correlated sets. Results demonstrated that the extracted sets of genes and diseases provided meaningful explanation to the molecular causes of the miRNA-disease associations, and that diseases in each correlated set were outcomes from miRNA perturbations of target genes. When applied to collected data sets for 5-fold cross-validation experiments, our method obtained an AUC value of 0.84606. We finally conducted comprehensive miRNA-disease association predictions and confirmed some high-ranking results using independent source of information.

Results

Extraction of canonical component sets of miRNA-targeted genes and miRNA-related diseases

The proposed method (see Methods) was applied to the target gene profiles and disease profiles to receive 60 canonical components (CC). The extracted genes and diseases in each component were available at Additional file 1. We also provided information of miRNAs that contributed to the correlations. It could be discovered that each component includes a small number of genes and diseases, which indicated an advantage of adding parameters c_1 and c_2 to impose sparseness on ordinary canonical correlation analysis (OCCA). It should be noted that experiments suggested that all the weight vectors received by OCCA were not sparse with

Table 1 The top 10 predicted results for *hsa-miR-203a*

miRNA	rank	predicted disease	confirmed
<i>hsa-miR-203a</i>	1	Colorectal Neoplasms	
<i>hsa-miR-203a</i>	2	Breast Neoplasms	Yes
<i>hsa-miR-203a</i>	3	Stomach Neoplasms	Yes
<i>hsa-miR-203a</i>	4	Melanoma	
<i>hsa-miR-203a</i>	5	Carcinoma, Hepatocellular	Yes
<i>hsa-miR-203a</i>	6	Lung Neoplasms	Yes
<i>hsa-miR-203a</i>	7	Prostate Neoplasms	Yes
<i>hsa-miR-203a</i>	8	Heart Failure	
<i>hsa-miR-203a</i>	9	Carcinoma, Non-Small-Cell Lung	
<i>hsa-miR-203a</i>	10	Urinary Bladder Neoplasms	

the parameter K ranged from 10 to 200 by 10 increments, which prevented the extraction of CCs.

Interpretation of miRNA-disease associations through the extracted component sets

It has been known that miRNAs exert their biological functions by regulating the expressions of their target genes. Our method was based on this conclusion and we intended to interpret the cause-and-effect mechanism behind the existing miRNA-disease associations based on the extracted component sets. We chose CC31, CC40 and CC54 in the extracted 60 canonical components as examples.

In CC31, we received 3 miRNAs (*hsa-miR-34a*, *hsa-miR-34b* and *hsa-miR-34c*). The numbers of target genes for the 3 miRNAs were 139, 32 and 50, respectively. The top-ranked target gene for the 3 miRNAs was *ZAP70*. Experimental evidence indicated that tyrosine kinase encoded by this gene plays an essential role in regulation, such as immune response, thymocyte development and cytokine expression of mature T-cells. Meanwhile, the top-predicted disease was *sarcoma*, which was a highly malignant tissue neoplasm caused by proliferation of mesodermal cells. Records in the latest version of

Table 2 The top 10 predicted diseases by our method

Rank	Our prediction
1	Carcinoma, Hepatocellular
2	Breast Neoplasms
3	Colorectal Neoplasms
4	Stomach Neoplasms
5	Lung Neoplasms
6	Melanoma
7	Urinary Bladder Neoplasms
8	Ovarian Neoplasms
9	Glioblastoma
10	Glioma

CTD [39], a publicly available database curated information about environmental factors affecting human health, indicated that *ZAP70* was one of disease genes for *sarcoma*.

Two miRNAs (*hsa-miR-106b* and *hsa-miR-93*) were included in CC40. For *hsa-miR-106b*, it targeted 47 genes and 34 genes were targeted by *hsa-miR-93*. The top predicted genes were *DAB2* and *PTENP1* and the top ranked diseases were *Lung Diseases* and *Ovary Syndrome*. In CTD, we discovered that both genes were related to *Lung Diseases*. Even though, no information was available in CTD about the roles of the two genes in *Ovary Syndrome*. Studies showed that *DAB2* was expressed in normal ovarian epithelial cells and the down-regulation of *DAB2* may lead to ovarian carcinogenesis [40].

In CC54, 5 genes (*ANAPC1*, *RPIA*, *IGF2BP2*, *CYP2J2* and *LRIG1*) and 1 disease (*Cataract*) were extracted as a correlated set. Cataracts often affect old people, causing blurry vision. It was reported that more than half of Americans either have a cataract or have had cataract surgery by the age of 80 (<https://www.nei.nih.gov/health/ataract/>). Therefore, detecting the genetic causes of cataracts is of great importance. The retrieval of information from CTD suggested that all the 5 target genes were genetic factors of the disease *Cataract*.

It should be pointed out that some records of disease genes in CTD might be inferred. Even though, the co-occurrence of disease genes extracted by our method and recorded in the CTD database reinforced the reality of such information because our knowledge of disease genes was not complete.

Performance evaluation

To evaluate the prediction ability of our approach, we conducted 5-fold cross validations on the benchmark datasets. We first randomly split the 404 miRNAs into five subsets of roughly equal size. Each subset is then used in turn as a test set and training is performed on the remaining four subsets. For the test set, the related disease relationships of miRNAs are removed and we calculate the prediction scores based on the weight vectors of the components extracted from the training set. We rank the predicted miRNA-disease associations according to the prediction scores. The higher a validated miRNA-disease association is ranked, the better the prediction performance is. To obtain robust results, we repeated the cross-validation experiments five times.

Given a threshold δ , if the score of a predicted result is greater than δ , it is considered as a positive sample. Otherwise, it is deemed as a negative sample. To obtain a receiver operating characteristic (ROC) curve, the true

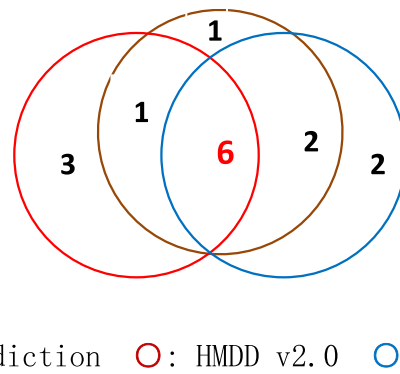


Fig. 1 Comparison of the top 10 diseases inferred by our method and the top 10 diseases with the highest values of MSW in HMDD v2.0 and HMDD v3.0

positive rates (TPRs) and the false positive rates (FPRs) at various δ values are computed as,

$$TPR = \frac{TP}{TP + FN}, FPR = \frac{FP}{TN + FP} \quad (1)$$

where TP and TN are the numbers of correctly identified positive and negative samples. FP and FN are the numbers of misidentified positive and negative samples. The area under the ROC curve (AUC) is used to measure the performance of our approach.

There are three parameters in our method. The parameters c_1 and c_2 are to control the sparseness, and K is the number of pairs of extracted canonical components. They are well tuned and we obtain the best AUC value of 0.84606 with $c_1 = 0.01$, $c_2 = 0.01$ and $K = 60$.

Prediction of novel miRNA-disease associations

After confirming the prediction performance by cross validations, we further applied our method to the miRNAs, which was not included in the benchmark datasets but whose target gene information was available, for their disease association predictions. There were 100 miRNAs of such kind in miRTarBase. We considered them as new miRNAs and their associated diseases needed to be predicted. The benchmark datasets were

taken as a training set. The predicted results for each new miRNA were listed in Additional file 2.

As we were preparing for the manuscript, the database HMDD v3.0 [41] was released. We took HMDD v3.0 as evidence to confirm the newly predicted miRNA-disease associations. We selected *hsa-miR-203a* as an example. For this miRNA, we chose the top 10 predicted results from the 362 candidate diseases and discovered that 5 of them were annotated by HMDD v3.0 (Table 1).

Furthermore, we chose the top 10 diseases (Table 2) predicted by our method and compared them with the top 10 diseases with the highest values of MSW (miRNA spectrum width) [41], an indicator of complexity of miRNA regulation, based on HMDD v2.0 and HMDD v3.0 (Fig. 1). As there are still associations between miRNAs and diseases undetected, the co-occurrence of top diseases in Fig. 1 demonstrates the inference capacity of our method from another way of perspective.

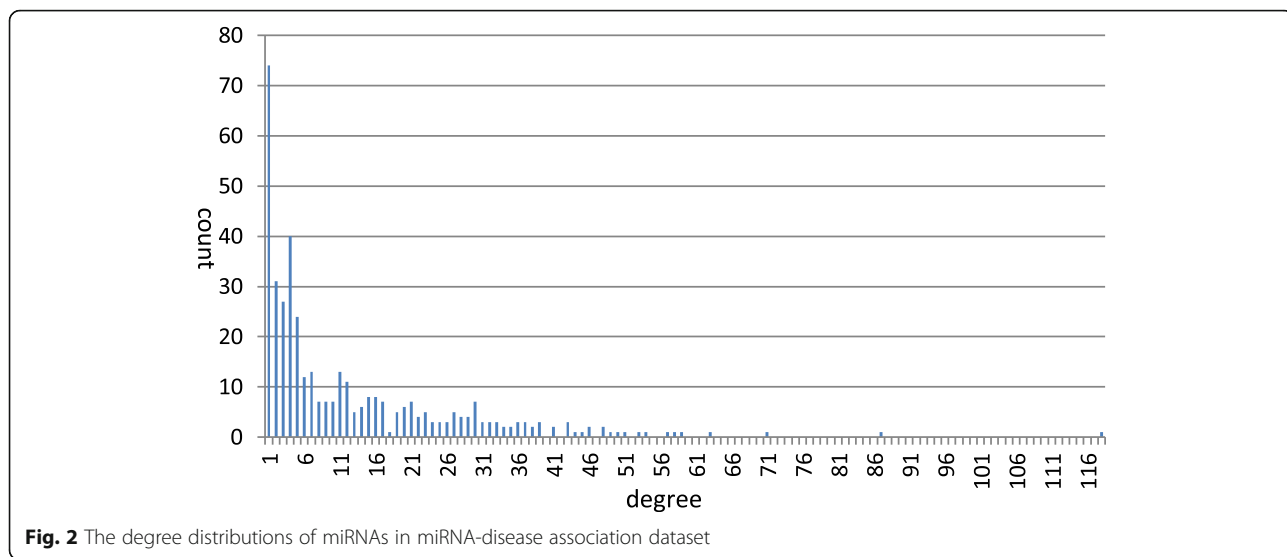
Discussion

Revealing the roles of miRNAs in diseases is critical for understanding the genetic causes of human diseases. Satisfied prediction performance could be obtained by previous methods for inferring miRNA-disease associations. However, they seldom investigated the underlying cause-and-effect mechanism involved. The correlated sets of target genes and human complex diseases received by our CCA-based method provided insights into the functional roles of miRNAs in the development of diseases. Three case studies also demonstrated correlation between the extracted target genes and diseases.

The excellent performance of our method could be contributed to two factors: data quality and additional penalty imposed on the CCA model. The benchmark datasets were received from two highly reliable databases and the miRNA-gene interactions and miRNA-disease associations were supported by published papers. Furthermore, adding more penalties on ordinary CCA

Table 3 Statistics of the datasets used in our manuscript

Name	Statistics
# miRNAs	404
# target genes	2796
# diseases	362
# miRNA-gene interactions	7999
# miRNA-disease associations	5117
average number of target genes for each miRNA	19.8
average number of related diseases for each miRNA	12.7

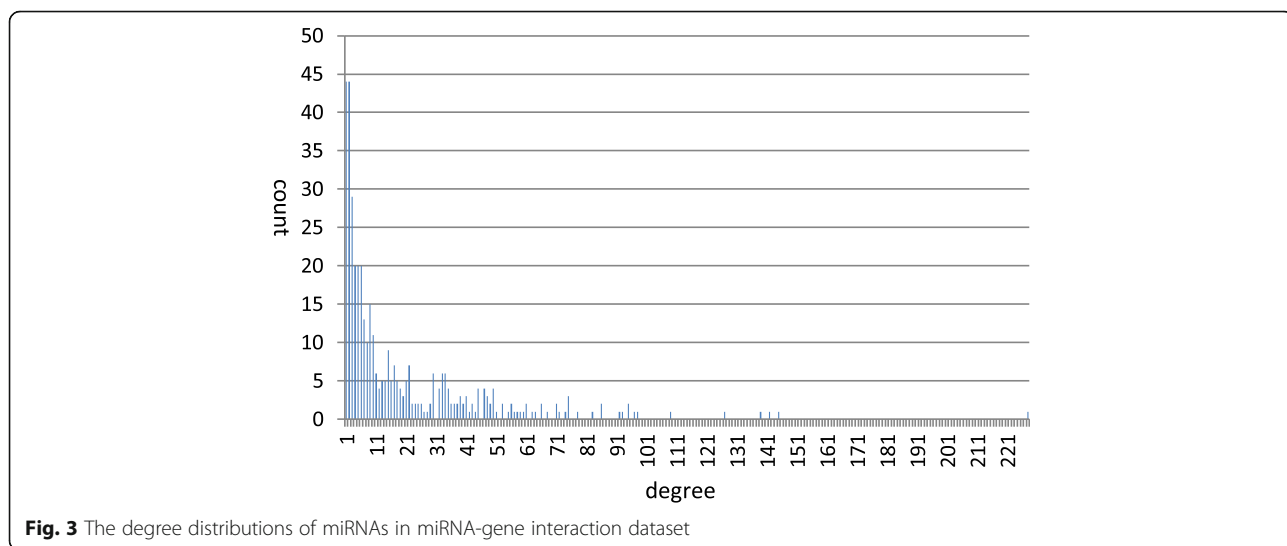


produced sparse weight vectors, which facilitate reasonable interpretation of the results.

As the data features used in our method were different from those in other computational models, a direct performance comparison was hard to implement. It is supposed in our method that a new miRNA's target gene information is available, but not disease information. Indeed, it is not practical in a real situation that the detailed target genes profile is known for a miRNA candidate molecule, which is a disadvantage in our method. However, advance in biological assays is bringing increasing information regarding miRNA-gene interactions. In this context, we expect that our proposed method presents itself as an informative tool for discovering the pathogenesis of diseases.

Conclusions

In this paper, we developed a novel method based on CCA with sparseness constraints for inferring and interpreting miRNA-disease associations. The results received from cross validations confirmed the excellent prediction power of our method. Experimental results also indicated that imposing the sparseness characteristic on CCA contributed to the improvement of interpretation of miRNA-disease associations. The extracted pairs of genes and diseases offered biological guidance for investigating how miRNA-disease associations were formed. When applying our method for predicting associated diseases for new miRNAs, some high-scoring results were supported by HMDD v3.0.



Input: the target gene profile X and the disease profile Y of miRNAs, parameters c_1, c_2 and K , the gene profile of a new miRNA x_{new}

Output: the predicted disease profile y_{new}

- 1 $Z^1 = X^T Y$
- 2 For $1 \leq k \leq K$
 apply PMD on Z^k to obtain weight vectors α_k and β_k , and singular value d_k .
 $Z^{k+1} = Z^k - d_k \alpha_k \beta_k^T$
- 3 End For
- 4 For a new miRNA, whose target gene profile is x_{new} , its disease profile is calculated as

$$y_{new} = \sum_{k=1}^K \beta_k \rho_k \alpha_k^T x_{new}$$
 (ρ_k is the canonical correlation coefficient received each iteration)

Fig. 4 Description of the algorithm SCCA

Methods

Data preparation

Two datasets, namely miRNA-gene interactions and miRNA-disease associations, were used in our method. We downloaded miRNA-gene interactions from miRTarBase [42], which was built by manually surveying pertinent literature to retrieve experimentally confirmed miRNA-gene interactions. For these interactions, we constrained the miRNA species into *Homo sapiens*. The interactions supported by weak experimental evidences were not taken into consideration. miRNA-disease associations were received from HMDD v2.0 [14] whose experimentally validated associations between miRNAs

and diseases were manually retrieved from literature. For both datasets, we merged the records of different miRNA copies that produce the same mature miRNA. Diseases with synonyms in HMDD v2.0 were also merged and invalid disease names were filtered out. Redundant records in the two datasets were kept only once.

In total, 404 miRNAs had both target gene and disease information. We finally received 7999 miRNA-gene interactions containing 2796 target genes and 5117 miRNA-disease pairs containing 362 diseases (Table 3). Intuitively, we could model each of the two datasets as a bipartite graph. The degree distributions of miRNAs,

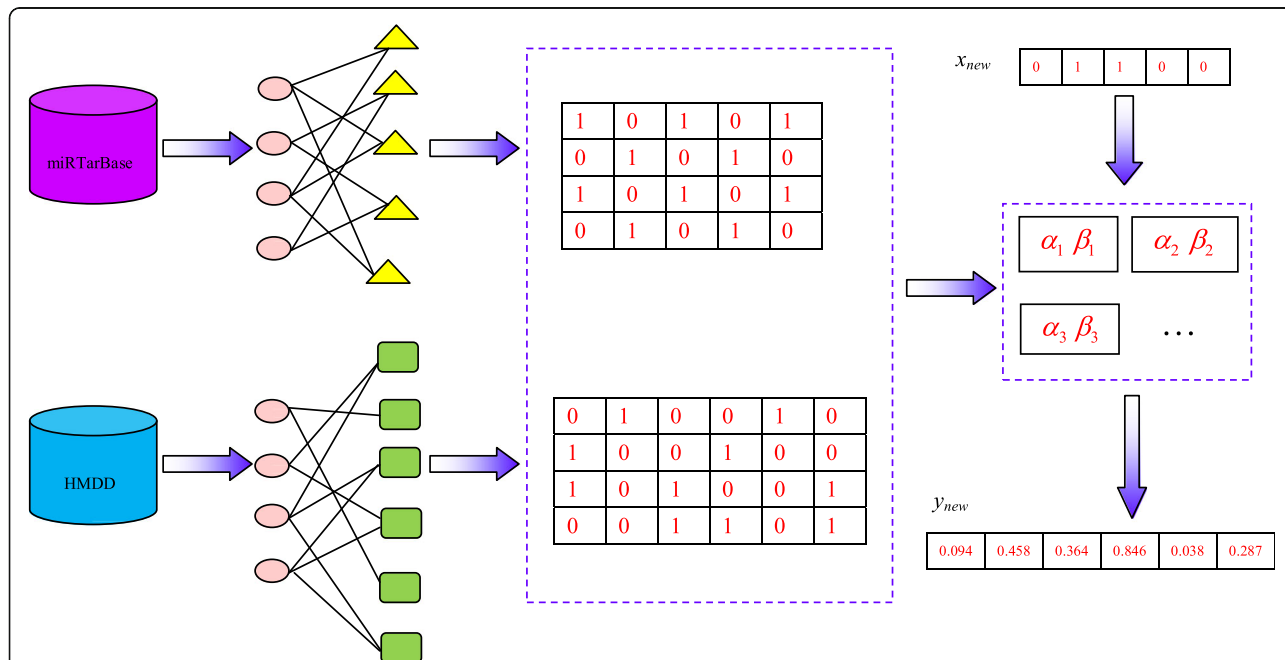


Fig. 5 Schematic of our proposed model. First, we extracted miRNA-gene interactions and miRNA-disease associations from miRTarBase and HMDD, respectively. Then, target gene profiles and disease profiles for miRNAs were constructed. Third, canonical correlation analysis was performed to obtain correlated sets. Finally, novel miRNA-disease associations were predicted based on the weight vectors of the correlated sets.

which indicate the numbers of target genes or the numbers of associated diseases of miRNAs, in each graph were listed in Fig. 2 and Fig. 3, respectively. The two sets of data were regarded as benchmark datasets and were used for performance evaluation in the following cross-validation experiments as well as training data for novel miRNA-disease association predictions.

Method description

Construction of gene profiles and disease profiles for miRNAs

Suppose that we have a set of n miRNAs which included p target genes and q related disease features. Each miRNA can then be denoted by a target gene profile $t = (t_1, t_2, t_3, \dots, t_p)^T$ and by a disease profile $d = (d_1, d_2, d_3, \dots, d_q)^T$, where t_i (or d_j) is represented for the presence or absence of gene (or disease) by 1 or 0, respectively.

Canonical correlation analysis (CCA)

Two linear combinations for target gene profiles and disease profiles are defined as $u_i = \alpha^T t_i$ and $v_i = \beta^T d_i$ ($i = 1, 2, 3, \dots, n$), where $\alpha = (\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_p)^T$ and $\beta = (\beta_1, \beta_2, \beta_3, \dots, \beta_q)^T$ are weight vectors. Our goal is to find weight vectors α and β which maximize the following canonical correlation coefficient:

$$\rho = \text{corr}(u, v) = \frac{\sum_{i=1}^n \alpha^T t_i \cdot \beta^T d_i}{\sqrt{\sum_{i=1}^n (\alpha^T t_i)^2} \sqrt{\sum_{i=1}^n (\beta^T d_i)^2}} \quad (2)$$

where u and v are centered, respectively.

Let X denote the $n \times p$ matrix and Y denote the $n \times q$ matrix. Then the maximization problem can be formally rewritten as follows:

$$\text{maximize} \{ \alpha^T X^T Y \beta \} \text{ subject to } \|\alpha\|_2^2 \leq 1, \|\beta\|_2^2 \leq 1. \quad (3)$$

Canonical correlation analysis (CCA), developed by Hotelling [43], provided a solution to the problem. We consider it as ordinary canonical correlation analysis (OCCA). However, OCCA usually results in vectors α and β that are not sparse. We are interested in finding a linear combination of the variables in X and Y that has large correlation but is also sparse in the variables used. We therefore choose to add penalties in (2) and reconsider the maximization problem as:

$$\begin{aligned} &\text{maximize} \{ \alpha^T X^T Y \beta \} \\ &\text{subject to } \|\alpha\|_2^2 \leq 1, \|\beta\|_2^2 \leq 1, \|\alpha\|_1 \leq c_1 \sqrt{p}, \|\beta\|_1 \leq c_2 \sqrt{q} \end{aligned} \quad (4)$$

where c_1 and c_2 are parameters to control the sparsity. We refer to this as sparse canonical correlation analysis

(SCCA). We applied a strategy of penalized matrix decomposition (PMD) [44] on the matrix $Z = X^T Y$ to obtain the weight vectors α and β .

To compute multiple canonical variates, we performed PMD on the following Z^k iteratively.

1. Let $Z^1 = X^T Y$
2. For $k \in 1, 2, \dots, K - 1$

$Z^{k+1} = Z^k - d_k \alpha_k \beta_k^T$ (α_k and β_k are the weight vectors, and d_k is singular value obtained in each step).

Similar to [45], we selected genes and diseases in the K pairs of weight vectors with the highest values as correlated sets.

Novel miRNA-disease association predictions

To predict the related disease profile y_{new} of a new miRNA with a known target gene profile x_{new} , we calculate the prediction scores based on the K pairs of weight vectors received above according to the following equation:

$$y_{new} = \sum_{k=1}^K \beta_k \rho_k \alpha_k^T x_{new} \quad (5)$$

The elements in y_{new} with the highest scores are chosen as the potential diseases for the new miRNA. We outlined the complete steps for inferring potential miRNA-disease associations in Fig. 4, and the workflow of our model was illustrated in Fig. 5.

Additional files

- Additional file 1** The extracted 60 canonical components. (TXT 25 kb)
- Additional file 2** The predicted scores of diseases for the 100 new miRNAs. (XLS 427 kb)

Abbreviations

CC: Canonical component; CCA: Canonical correlation analysis; miRNAs: microRNAs; MSW: MiRNA spectrum width; PMD: Penalized matrix decomposition

Acknowledgements

We are grateful to Dr. Sayaka Mizutani at Tokyo Institute of Technology for useful discussions.

Authors' contributions

HC collected the data. HC, ZZ and DF conceived and designed the experiments. HC implemented the experiments. HC and ZZ analyzed the results. HC and ZZ wrote the paper. All authors read and approved the final manuscript

Funding

This work was supported by the National Natural Science Foundation of China under Grant 61862026 and by the Natural Science Foundation of Jiangxi, China, under Grant 20181BAB202008. The funding bodies did not play any roles in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹School of Software, East China Jiaotong University, Nanchang 330013, China.

²School of Computer Science and Engineering, Central South University, Changsha 410083, China.

Received: 4 May 2019 Accepted: 16 July 2019

Published online: 25 July 2019

References

- Lee RC, Feinbaum RL, Ambros V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *CELL*. 1993; 75(5):843–54.
- Kozomara A, Birgaoanu M, Griffiths-Jones S. miRBase: from microRNA sequences to function. *Nucleic Acids Res*. 2018;47:D155–D162.
- Hwang H, Mendell JT. MicroRNAs in cell proliferation, cell death, and tumorigenesis. *Br J Cancer*. 2006;94(6):776–80.
- Karp X, Ambros V. Encountering MicroRNAs in cell fate signaling. *Science*. 2005;310(5752):1288–9.
- Chen C, Li L, Lodish HF, Bartel DP. MicroRNAs modulate hematopoietic lineage differentiation. *Science*. 2004;303(5654):83–6.
- Shivdasani RA. MicroRNAs: regulators of gene expression and cell differentiation. *Blood*. 2006;108(12):3646–53.
- Wienholds E, Plasterk RHA. MicroRNA function in animal development. *FEBS Lett*. 2005;579(26):5911–22.
- Nagaraja AK, Creighton CJ, Yu Z, Zhu H, Gunaratne PH, Reid JG, Olokpa E, Itamochi H, Ueno NT, Hawkins SM, et al. A link between *mir-100* and *FRAP1/mTOR* in clear cell ovarian Cancer. *Mol Endocrinol*. 2010;24(2):447–63.
- Matsui M, Corey DR. Non-coding RNAs as drug targets. *Nat Rev Drug Discov*. 2016;16:167.
- Chen H, Zhang Z. A miRNA-driven inference model to construct potential drug-disease associations for drug repositioning. *Biomed Res Int*. 2015;2015:1–9.
- Chen H, Zhang Z, Peng W. miRDDCR: a miRNA-based method to comprehensively infer drug-disease causal relationships. *Sci Rep*. 2017;7(1):15921.
- Chen H, Zhang Z. Prediction of drug-disease associations for drug repositioning through drug-miRNA-disease heterogeneous network. *IEEE Access*. 2018;6:45281–7.
- Tang W, Wan S, Yang Z, Teschendorff AE, Zou Q. Tumor origin detection with tissue-specific miRNA and DNA methylation markers. *Bioinformatics*. 2018;34(3):398–406.
- 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*. 2013;42(D1):D1070–4.
- Xie B, Ding Q, Han H, Wu D. miRCancer: a microRNA-cancer association database constructed by text mining on literature. *Bioinformatics*. 2013; 29(5):638–44.
- Jiang Q, Wang Y, Hao Y, Juan L, Teng M, Zhang X, Li M, Wang G, Liu Y. miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res*. 2009;37(Database):D98–D104.
- Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*. 2004;116(2):281–97.
- Mork S, Pletscher-Frankild S, Palleja Caro A, Gorodkin J, Jensen LJ. Protein-driven inference of miRNA-disease associations. *Bioinformatics*. 2014;30(3):392–7.
- Li X, Wang Q, Zheng Y, Lv S, Ning S, Sun J, Huang T, Zheng Q, Ren H, Xu J, et al. Prioritizing human cancer microRNAs based on genes' functional consistency between microRNA and cancer. *Nucleic Acids Res*. 2011;39(22):e153.
- Shi H, Xu J, Zhang G, Xu L, Li C, Wang L, Zhao Z, Jiang W, Guo Z, Li X. Walking the interactome to identify human miRNA-disease associations through the functional link between miRNA targets and disease genes. *BMC Syst Biol*. 2013;7(1):101.
- Lu M, Zhang Q, Deng M, Miao J, Guo Y, Gao W, Cui Q. An analysis of human microRNA and disease associations. *PLoS One*. 2008;3(10):e3420.
- Chen X, Liu M, Yan G. RWRMDA: predicting novel human microRNA-disease associations. *Mol BioSyst*. 2012;8(10):2792.
- Chen H, Zhang Z. Prediction of associations between OMIM diseases and MicroRNAs by random walk on OMIM disease similarity network. *Sci World J*. 2013;2013:1–6.
- Chen H, Zhang Z. Similarity-based methods for potential human microRNA-disease association prediction. *BMC Med Genet*. 2013;6(1):12.
- Chen X, Yin J, Qu J, Huang L. MDHGI: matrix decomposition and heterogeneous graph inference for miRNA-disease association prediction. *PLoS Comput Biol*. 2018;14(8):e1006418.
- Liao B, Ding S, Chen H, Li Z, Cai L. Identifying human microRNA-disease associations by a new diffusion-based method. *J Bioinform Comput Biol*. 2015;13(04):1550014.
- Chen X, Huang L. LRSSLMDA: Laplacian regularized sparse subspace learning for miRNA-disease association prediction. *PLoS Comput Biol*. 2017; 13(12):e1005912.
- Lan W, Wang J, Li M, Liu J, Wu F, Pan Y. Predicting MicroRNA-disease associations based on improved MicroRNA and disease similarities. *IEEE/ACM Trans Comput Biol Bioinform*. 2018;15(6):1774–82.
- Zou Q, Li J, Hong Q, Lin Z, Wu Y, Shi H, Ju Y. Prediction of MicroRNA-disease associations based on social network analysis methods. *Biomed Res Int*. 2015;2015:1–9.
- Chen X, Yan G. Semi-supervised learning for potential human microRNA-disease associations inference. *Sci Rep UK*. 2014;4(1):5501.
- Chen X, Gong Y, Zhang DH, You ZH, Li ZW. DRMDA: deep representations-based miRNA-disease association prediction. *J Cell Mol Med*. 2018;22(1):472–85.
- Chen X, Wu Q, Yan G. RKNNDMA: ranking-based KNN for miRNA-disease association prediction. *RNA Biol*. 2017;14(7):952–62.
- Peng W, Lan W, Yu Z, Wang J, Pan Y. A framework for integrating multiple biological networks to predict MicroRNA-disease associations. *IEEE Trans Nanobioscience*. 2017;16(2):100–7.
- Xiao Q, Luo J, Liang C, Cai J, Ding P. A graph regularized non-negative matrix factorization method for identifying microRNA-disease associations. *Bioinformatics*. 2018;34(2):239–48.
- Liang C, Yu S, Luo J. Adaptive multi-view multi-label learning for identifying disease-associated candidate miRNAs. *PLoS Comput Biol*. 2019;15(4):e1006931.
- Li G, Luo J, Xiao Q, Liang C, Ding P. Predicting microRNA-disease associations using label propagation based on linear neighborhood similarity. *J Biomed Inform*. 2018;82:169–77.
- Liu Y, Zeng X, He Z, Zou Q. Inferring MicroRNA-disease associations by random walk on a heterogeneous network with multiple data sources. *IEEE/ACM Trans Comput Biol Bioinform*. 2017;14(4):905–15.
- Zeng X, Liu L, Lü L, Zou Q. Prediction of potential disease-associated microRNAs using structural perturbation method. *Bioinformatics*. 2018;34(14):2425–32.
- Davis AP, Grondin CJ, Johnson RJ, Sciaky D, McMorran R, Wieggers J, Wieggers TC, Mattingly CJ. The comparative Toxicogenomics database: update 2019. *Nucleic Acids Res*. 2019;47(D1):D948–54.
- Mok SC, Chan WY, Wong KK, Cheung KK, Lau CC, Ng SW, Baldini A, Colitti CV, Rock CO, Berkowitz RS. DOC-2, a candidate tumor suppressor gene in human epithelial ovarian cancer. *Oncogene*. 1998;16(18):2381–7.
- Huang Z, Shi J, Gao Y, Cui C, Zhang S, Li J, Zhou Y, Cui Q. HMDD v3.0: a database for experimentally supported human microRNA-disease associations. *Nucleic Acids Res*. 2019;47(D1):D1013–7.
- Chou CH, Shrestha S, Yang CD, Chang NW, Lin YL, Liao KW, Huang WC, Sun TH, Tu SJ, Lee WH, et al. miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic Acids Res*. 2018;46(D1):D296–302.
- Hotelling H. Relations between two sets of variates. *Biometrika*. 1936;28:321–77.
- Witten DM, Tibshirani R, Hastie T. A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics*. 2009;10(3):515–34.
- Mizutani S, Pauwels E, Stoven V, Goto S, Yamanishi Y. Relating drug-protein interaction network with drug side effects. *Bioinformatics*. 2012;28(18):i522–8.

Publisher's Note

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