BMC Bioinformatics

Open Access

Prediction of microRNA-disease associations based on distance correlation set

Haochen Zhao[2,5], Linai Kuang[1,2,5], Lei Wang[1,2,3,5*], Pengyao Ping[2,5], Zhanwei Xuan[2,5], Tingrui Pei[2,5] and Zhelun Wu[4]

Abstract

Background: Recently, numerous laboratory studies have indicated that many microRNAs (miRNAs) are involved in and associated with human diseases and can serve as potential biomarkers and drug targets. Therefore, developing effective computational models for the prediction of novel associations between diseases and miRNAs could be beneficial for achieving an understanding of disease mechanisms at the miRNA level and the interactions between diseases and miRNAs at the disease level. Thus far, only a few miRNA-disease association pairs are known, and models analyzing miRNA-disease associations based on lncRNA are limited.

Results: In this study, a new computational method based on a distance correlation set is developed to predict miRNA-disease associations (DCSMDA) by integrating known lncRNA-disease associations, known miRNA-lncRNA associations, disease semantic similarity, and various lncRNA and disease similarity measures. The novelty of DCSMDA is due to the construction of a miRNA-lncRNA-disease network, which reveals that DCSMDA can be applied to predict potential lncRNA-disease associations without requiring any known miRNA-disease associations. Although the implementation of DCSMDA does not require known disease-miRNA associations, the area under curve is 0.8155 in the leave-one-out cross validation. Furthermore, DCSMDA was implemented in case studies of prostatic neoplasms, lung neoplasms and leukaemia, and of the top 10 predicted associations, 10, 9 and 9 associations, respectively, were separately verified in other independent studies and biological experimental studies. In addition, 10 of the 10 (100%) associations predicted by DCSMDA were supported by recent bioinformatical studies.

Conclusions: According to the simulation results, DCSMDA can be a great addition to the biomedical research field.

Keywords: MiRNA-disease association predictions, Distance correlation set, Disease-lncRNA-miRNA network, Similarity measure

Background

For a long time, RNA was considered a DNA-to-protein gene sequence transporter [1]. The sequencing of the human genome indicates that only approximately 2% of the sequences in human RNA are used to encode proteins [2]. Furthermore, numerous studies performing biological experiments have indicated that noncoding RNA (ncRNA) plays an important role in numerous critical biological processes, such as chromosome dosage compensation, epigenetic regulation and cell growth [3–5]. MicroRNAs (miRNAs) are endogenous single-stranded ncRNA molecules approximately 22 nt in length that regulate the expression of target genes by base pairing with the 3′-untranslated regions (UTRs) of the target genes [6, 7]. Recently, several studies have reported that more than one-third of genes are regulated by miRNAs [8], and more than 1000 miRNAs have been identified using various experimental methods and computational models [9, 10]. In addition, accumulating evidence indicates that many microRNAs (miRNAs) are involved in and associated with human diseases, such as myocardial disease, Alzheimer's disease, cardiovascular disease and heart disease [11–14]. Therefore, identifying disease-

* Correspondence: wanglei@xtu.edu.cn
[1]College of Computer Engineering & Applied Mathematics, Changsha University, Changsha 410001, Hunan, People's Republic of China
[2]Key Laboratory of Intelligent Computing & Information Processing (Xiangtan University), Ministry of Education, China, Xiangtan 411105, Hunan, People's Republic of China
Full list of author information is available at the end of the article

Zhao *et al. BMC Bioinformatics* (2018) 19:141

Page 2 of 14

miRNA associations could not only improve our knowledge of the underlying disease mechanism at the miRNA level but also facilitate disease biomarker detection and drug discovery for disease diagnosis, treatment, prognosis and prevention. However, compared with the rapidly increasing number of newly discovered miRNAs, only a few miRNA-disease associations are known [15, 16]. Developing efficient, successful computational approaches that predict potential miRNA-disease associations is challenging and urgently needed.

Recently, several heterogeneous biological datasets, such as HMDD and miR2Disease, have been constructed [17–19], and several computational methods are used to predict potential miRNA-disease associations based these datasets [20–22]. For example, Jiang et al. developed a scoring system to assess the likelihood that a microRNA is involved in a specific disease phenotype based on the assumption that functionally related microRNAs tend to be associated with phenotypically similar diseases [23]. K. Han et al. developed a prediction method called DismiPred that combines functional similarity and common association information to predict potential miRNA-disease associations based on the central hypothesis offered in several previous studies that miRNAs with similar functions are often involved in similar diseases [24]. Furthermore, Xuan et al. proposed a method called HDMP to predict potential disease-miRNA associations based on weighted k most similar neighbours [25] and developed a method for predicting potential disease-associated microRNAs based on random walk (MIDP) [26]. Chen et al. proposed a prediction method called RWRMDA by implementing random walk on the miRNA functional similarity network and further proposed a model called RLSMDA based on semi-supervised learning by integrating a disease-disease semantic similarity network, miRNA-miRNA functional similarity network, and known human miRNA-disease associations for the prediction of potential disease-miRNA associations [27]. In 2016, based on the assumption that functionally similar miRNAs tend to be involved in similar diseases, Chen et al. developed a prediction model called WBSMDA by integrating known miRNA-disease associations, miRNA functional similarity networks, disease semantic similarity networks, and Gaussian interaction profile kernel similarity networks to uncover potential disease-miRNA associations [28].

In the abovementioned computational models, known miRNA-disease associations are required. However, few lncRNA-disease associations have been recorded in several biological datasets, such as MNDR and LncRNADisease [29, 30], and several studies have shown that lncRNA-miRNA associations are involved in and associated with human diseases [31–33]. Thus, in this article, a new model based on the Distance Correlation Set for MiRNA-Disease

Association inference (DCSMDA) was developed to predict potential miRNA-disease associations by integrating known lncRNA-disease and lncRNA-miRNA associations, the semantic similarity and functional similarity of the disease pairs, the functional similarity of the miRNA pairs, and the Gaussian interaction profile kernel similarity for the lncRNA, miRNA and disease. Compared with existing state-of-the-art models, the advantage of DCSMDA is its integration of the similarity of the disease pairs, lncRNA pairs, miRNA pairs, and introduction of the distance correlation set; thus, DCSMDA does not require known miRNA-disease associations. Moreover, leave-one-out cross-validation (LOOCV) was implemented to evaluate the performance of DCSMDA based on known miRNA-disease associations downloaded from the HMDD database, and DCSMDA achieved a reliable area under the ROC curve (AUC) of 0.8155. Moreover, case studies of lung neoplasms, prostatic neoplasms and leukaemia were implemented to further evaluate the prediction performance of DCSMDA, and 9, 10 and 9 of the top 10 predicted associations in these three important human complex diseases have been confirmed by recent biological experiments. In addition, a case study identifying the top 10 lncRNA-disease associations showed that 10 of the 10 (100%) associations predicted by DCSMDA were supported by recent bioinformatical studies and the latest HMDD dataset, effectively demonstrating that DCSMDA had a good prediction performance in inferring potential disease-miRNA associations.

Results

To evaluate the prediction performance of DCSMDA, first, our method was compared with other state-of-the-art methods in the framework of the LOOCV, and then, we analyzed the stability of DCSMDA using three lncRNA-disease datasets. Second, we analyzed the effect of the pre-determined threshold parameter b. Finally, several additional experiments were performed to validate the feasibility of our method.

Performance comparison with other methods

Since our method is unsupervised (i.e., known miRNA-disease associations are not used in the training) and the few proposed prediction models for the large-scale forecasting of the associations between miRNAs and diseases are simultaneously based on known miRNA-lncRNA associations and known lncRNA-disease associations, to validate the prediction performance of our novel model, we compared the prediction performance of DCSMDA with that of three state-of-the-art computational prediction models, including WBSMDA [28], RLSMDA [27] and HGLDA [31]; WBSMDA and RLSMDA are semi-supervised methods that do not require any negative samples, and HGLDA is an unsupervised method developed to predict potential

Zhao *et al. BMC Bioinformatics* (2018) 19:141

Page 3 of 14

lncRNA-disease associations by integrating known miRNA-disease associations and lncRNA-miRNA interactions.

To compare the performance of DCSMDA with that of WBSMDA and RLSMDA, we adopted the DS_5 dataset and the framework of the LOOCV. While the LOOCV was implemented for these three methods, each known miRNA-disease association was left out in turn as the test sample, and we further evaluated how well this test association ranked relative to the candidate sample. Here, the candidate samples comprised all potential miRNA-disease associations without any known association evidence. Then, the testing samples with a prediction rank higher than the given threshold were considered successfully predicted. If the testing samples with a prediction rank higher than the given threshold were considered successfully predicted, then DCSMDA, RLSMDA and WBSMDA were checked in the LOOCV.

To compare the performance of DCSMDA with that of HGLDA, we adopted the DS_3 dataset and the framework of the LOOCV. While the LOOCV was implemented for HGLDA, each known lncRNA-disease association was removed individually as a testing sample, and we further evaluated how well this test lncRNA-disease association ranked relative to the candidate sample. Here, the candidate samples comprised all potential lncRNA-disease associations without any known association evidence.

Thus, we could further obtain the corresponding true positive rates (TPR, sensitivity) and false positive rates (FPR, 1-specificity) by setting different thresholds. Here, sensitivity refers to the percentage of test samples that were predicted with ranks higher than the given threshold, and the specificity was computed as the percentage of negative samples with ranks lower than the threshold.

The receiver-operating characteristic (ROC) curves were generated by plotting the TPR versus the FPR at different thresholds. Then, the AUCs were further calculated to evaluate the prediction performance of DCSMDA.

An AUC value of 1 represented a perfect prediction, while an AUC value of 0.5 indicated a purely random performance. The performance comparison in terms of the LOOCV results is shown in Fig. 1. In the LOOCV, the DCSMDA (when b was set to 6), RLSMDA, WBSMDA and HGLDA achieved AUCs of 0.8155, 0.7826, 0.7582 and 0.7621, respectively. DCSMDA predicted potential miRNA-disease associations without requiring known miRNA-disease associations. To the best of our knowledge, no methods that rely on known miRNA-disease associations exist. More importantly, considering that known disease-lncRNA associations remain very limited, the performance of DCSMDA can be further improved as additional known miRNA-disease associations are obtained in the future.

The stability analysis of DCSMDA

Because the current lncRNA-disease databases remain in their infancy and most existing methods are always evaluated using a specific dataset, the stability of the different datasets is ignored. To enhance the credibility of the prediction results, DCSMDA was further implemented using three different known lncRNA-disease association datasets, including DS_1, DS_2, and DS_3, and the known lncRNA-miRNA association dataset DS_4.

The comparison results of the ROC are shown in Fig. 2, and the corresponding AUCs are 0.8155, 0.8089 and 0.7642 when DCSMDA (b was set to 6) was evaluated in the framework of the LOOCV using the three different
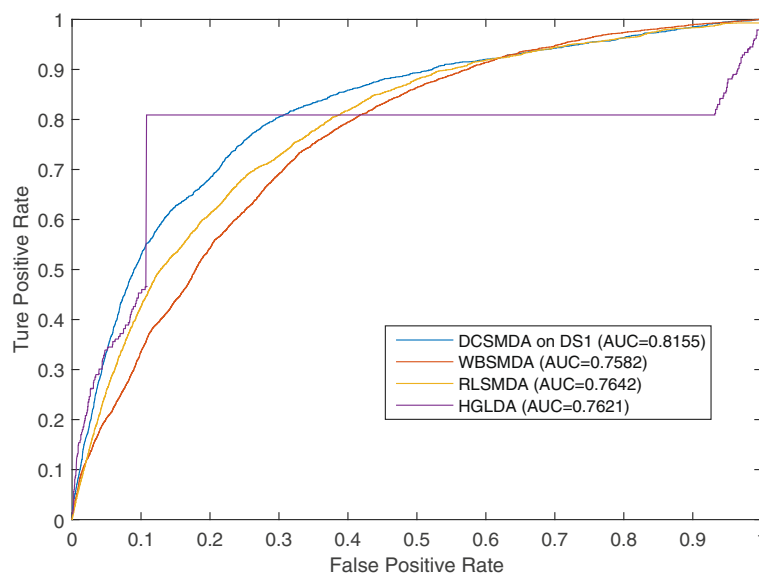
Fig. 1 Performance comparisons between DCSMDA, RLSMDA and HGLDA in terms of ROC curve and AUC based on LOOCV

Zhao *et al. BMC Bioinformatics* (2018) 19:141

Page 4 of 14
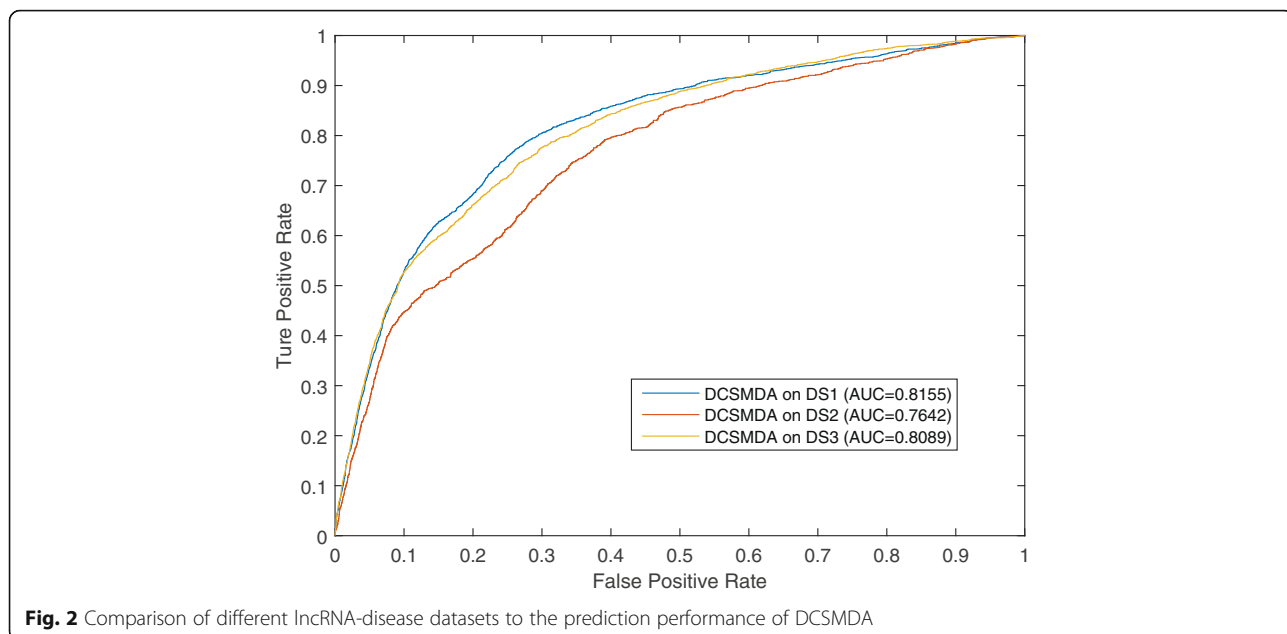
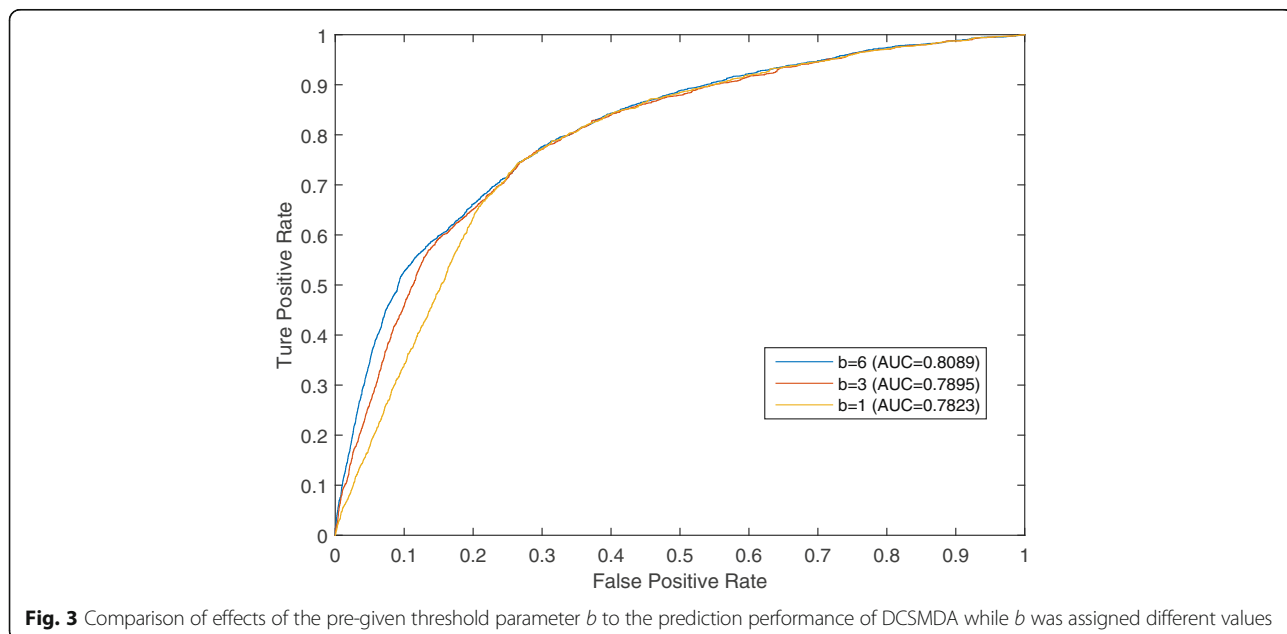**Fig. 2** Comparison of different lncRNA-disease datasets to the prediction performance of DCSMDA

lncRNA-disease association datasets. DCSMDA achieved a reliable and effective prediction performance.

Effects of the pre-given threshold parameter b

In DCSMDA, the pre-determined threshold b plays a critical role, and the value of b influences the performance of predicting potential miRNA-disease associations. In this section, we implemented a series of comparison experiments to evaluate the effects of b on the prediction performance of DCSMDA. The LOOCV was implemented, experiments were performed, and b was assigned different values. Considering the time complexity, and that the

value of SPM(i, j) always equals 6, when $b \geq 6$, we set b to a value no greater than 6 in our experiments.

As shown in Fig. 3, DCSMDA showed an increasing trend in its prediction performance as the value of the pre-determined threshold parameter b increased and achieved the best prediction performance when b was set to 6. When b was set to 6, DCSMDA achieved an AUC of 0.8089 using DS_3 and DS_4. In the analysis, we found that the main reason was that the number of known miRNA-lncRNA associations and lncRNA-disease associations was small; thus, when b is set to a larger value, more nodes could be linked to each other in the miRNA-

Fig. 3 Comparison of effects of the pre-given threshold parameter b to the prediction performance of DCSMDA while b was assigned different values

Zhao *et al. BMC Bioinformatics* (2018) 19:141

Page 5 of 14

lncRNA-disease interactive network, improving the prediction performance of DCSMDA. Therefore, we finally set $b = 6$ in our experiments.

Case study

Currently, cancer is the leading cause of death in humans worldwide [34–36], and the incidence of cancer is high in both developed and developing countries. Therefore, to estimate the effective predictive performance of DCSMDA, case studies of two important cancers and leukaemia were implemented. The prediction results were verified using recently published experimental studies (see Table 1).

Table 1 DCSMDA was applied to case studies of three important cancers. In total, 10, 9 and 8 of the top 10 predicted pairs for these diseases were confirmed based on recent experimental studies

Disease	miRNA	Evidence (PMID and PMCID)
Prostatic Neoplasms	hsa-mir-15a	PMID: 25418933
Prostatic Neoplasms	hsa-mir-15b	PMID: 24661838
Prostatic Neoplasms	hsa-mir-16	PMID: 21880514
Prostatic Neoplasms	hsa-mir-195	PMID: 26080838
Prostatic Neoplasms	hsa-mir-424	PMID: 27820701
Prostatic Neoplasms	hsa-mir-497	PMID: 23886135
Prostatic Neoplasms	hsa-mir-125a	PMCID: PMC3979818
Prostatic Neoplasms	hsa-mir-106b	PMID: 26124181
Prostatic Neoplasms	hsa-mir-17	PMCID: PMC3008681
Prostatic Neoplasms	hsa-mir-93	PMID: 26124181
Lung Neoplasms	hsa-mir-15a	PMID: 26314859
Lung Neoplasms	hsa-mir-195	PMID: 25840419
Lung Neoplasms	hsa-mir-424	PMID: 27666545
Lung Neoplasms	hsa-mir-497	PMCID: PMC4537005
Lung Neoplasms	hsa-mir-16	PMID: 21192009
Lung Neoplasms	hsa-mir-15b	Unconfirmed
Lung Neoplasms	hsa-mir-125a	PMID: 24044511
Lung Neoplasms	hsa-mir-106a	PMID: 18328430
Lung Neoplasms	hsa-mir-106b	PMID: 18328430
Lung Neoplasms	hsa-mir-93	PMID: 24037530
Leukaemia	hsa-mir-424	PMID: 27013583
Leukaemia	hsa-mir-195	PMCID: PMC4713510
Leukaemia	hsa-mir-16	PMID:22912766
Leukaemia	hsa-mir-15a	PMID: 24392455
Leukaemia	hsa-mir-15b	PMCID: PMC4577143
Leukaemia	hsa-mir-497	Unconfirmed
Leukaemia	hsa-mir-125a	PMID: 22456625
Leukaemia	hsa-mir-19b	PMID: 28765931
Leukaemia	hsa-mir-19a	PMID: 28765931
Leukaemia	hsa-mir-17	PMID: 20439436

Prostate cancer (prostatic neoplasms), which is the second leading cause of cancer-related death in males, is among the most common malignant cancers and the most commonly diagnosed cancer in men worldwide. In 2012, prostate cancer occurred in 1.1 million men and caused 307,000 deaths. Accumulating evidence shows that microRNAs are strongly associated with prostate cancer. Therefore, DCSMDA was implemented to predict potential prostate cancer-related miRNAs. Consequently, ten of the top ten predicted prostate cancer-related miRNAs were validated by recent biological experimental studies (see Table 1). For example, Junfeng Jiang et al. reconstructed five prostate cancer co-expressed modules using functional gene sets defined by Gene Ontology (GO) annotation (biological process, GO_BP) and found that hsa-mir15a (ranked 1st) regulated these five candidate modules [37]. Medina-Villaamil V et al. analyzed circulating miRNAs in whole blood as non-invasive markers in patients with localized prostate cancer and healthy individuals and found that hsa-mir-15b (ranked 2nd) showed a statistically significant differential expression between the different risk groups and healthy controls [38]. Furthermore, Chao Cai et al. confirmed the tumour suppressive role of hsa-mir-195 (ranked 4th) using prostate cancer cell invasion, migration and apoptosis assays in vitro and tumour xenograft growth, angiogenesis and invasion in vivo by performing both gain-of-function and loss-of-function experiments [39].

Lung cancer (lung neoplasms) has the poorest prognosis among cancers and is the largest threat to people's health and life. The incidence and mortality of lung cancer are rapidly increasing in China, and approximately 1. 4 million deaths are due to lung cancer annually. Recent studies show that miRNAs play critical roles in the progression of lung cancer. Therefore, we used lung cancer as a case study and implemented DCSMDA; nine predicted lung cancer-associated miRNAs of the top ten prediction list were verified based on experimental reports. For example, Bozok Çetintaş V et al. analyzed the effects of selected miRNAs on the development of cisplatin resistance and found that hsa-mir-15a (ranked 1st) was among the most significantly downregulated miRNAs conferring resistance to cisplatin in Calu1 epidermoid lung carcinoma cells [40]. Hsa-mir-195, which ranked 2nd, was further confirmed to suppress tumour growth and was associated with better survival outcomes in several malignancies, including lung cancer [41]. Additionally, according to the biological experiments reported in several studies, hsa-mir-424 (ranked 3rd) plays an important role in lung cancer [42].

Leukaemia refers to a group of diseases that usually begin in the bone marrow and result in high numbers of abnormal white blood cells. The exact cause of leukaemia is unknown, and a combination of genetic factors

Zhao *et al. BMC Bioinformatics* (2018) 19:141

Page 6 of 14

and environmental factors is believed to play a role. In 2015, leukaemia presented in 2.3 million people and caused 353,500 deaths. Several studies suggest that miRNAs are effective prognostic biomarkers in leukaemia. For example, independent experimental observations showed relatively lower expression levels of mir-424 (ranked 1st) in TRAIL-resistant and semi-resistant acute myeloid leukaemia (AML) cell lines and newly diagnosed patient samples. The overexpression of mir-424 by targeting the 3′ UTR of PLAG1 enhanced TRAIL sensitivity in AML cells [43]. Hsa-mir-16 ranked 3rd, its expression was inversely correlated with Bcl2 expression in leukaemia, and both microRNAs negatively regulate B cell lymphoma 2 (Bcl2) at a posttranscriptional level. Bcl2 repression by these microRNAs induces apoptosis in a leukaemic cell line model [44]. The lncRNA H19 is considered an independent prognostic marker in patients with tumours. The expression of lncRNA H19 is significantly upregulated in bone marrow samples from patients with AML-M2. The results of the current study suggest that lncRNA H19 regulates the expression of inhibitor of DNA binding 2 (ID2) by competitively binding to hsa-mir-19b (ranked 8) and hsa-mir-19a (ranked 9), which may play a role in AML cell proliferation [45].

In addition, DCSMDA predicted all potential associations between the diseases and miRNAs in G_3 simultaneously. In addition, notably, potential associations with a high predicted value can be publicly released and benefit from biological experimental validation. To further illustrate the effective performance of DCSMDA, the predicted results were sorted from best to worse, and the top 10 results were selected for analysis (see Table 2). Consequently, 100% of the results were confirmed by recent biological experiments and the HMDD dataset, and thus, DCSMDA can be used as an efficient computational tool in biomedical research studies.

Table 2 The top 10 predicted miRNA-disease associations by DCSMDA

Disease	MiRNA	Evidence
Carcinoma, Hepatocellular	hsa-mir-15a	HMDD
Carcinoma, Hepatocellular	hsa-mir-15b	HMDD
Carcinoma, Hepatocellular	hsa-mir-16	HMDD
Carcinoma, Hepatocellular	hsa-mir-195	HMDD
Carcinoma, Hepatocellular	hsa-mir-424	PMID: 26823812
Carcinoma, Hepatocellular	hsa-mir-497	HMDD
Colorectal Neoplasms	hsa-mir-497	HMDD
Colorectal Neoplasms	hsa-mir-15b	PMID: 23267864
Colorectal Neoplasms	hsa-mir-16	HMDD
Colorectal Neoplasms	hsa-mir-195	HMDD

Discussion

Accumulating evidence shows that miRNAs play a very important role in several key biological functions and signalling pathways. A large-scale systematic analysis of miRNA-disease data performed by combining relevant biological data is highly important for humans and attractive topics in the field of computational biology. However, only a few prediction models have been proposed for the large-scale forecasting of associations between miRNAs and diseases based on lncRNA information. To utilize the wealth of disease-lncRNA, miRNA-lncRNA and disease-lncRNA association data recorded in four datasets and recently published experimental studies, in this article, we proposed a novel prediction model called DCSMDA to infer the potential associations between diseases and miRNAs. We first constructed a miRNA-lncRNA-disease interactive network and further integrated a distance correlation set, disease semantic similarity, functional similarity and Gaussian interaction profile kernel similarity for DCSMDA. The important difference between DCSMDA and previous computational models is that DCSMDA does not rely on any known miRNA-disease associations and predicts disease-miRNA associations based only on known disease-lncRNA associations and known lncRNA-miRNA associations. To evaluate the prediction performance of DCSMDA, the validation frameworks of the LOOCV were implemented using the HMDD database. Furthermore, case studies were further implemented using three important diseases and the top 10 predicted miRNA-disease associations based on recently published experimental studies and databases. The simulation results showed that DCSMDA achieved a reliable and effective prediction performance. Hence, DCSMDA could be used as an effective and important biological tool that benefits the early diagnosis and treatment of diseases and improves human health in the future.

However, although DCSMDA is a powerful method for predicting novel relationships between diseases and miRNAs, there are several limitations in our method. First, the value of the threshold parameter b plays an important role in DCSMDA, and the selection of a suitable value for b is a critical problem that should be addressed in future studies. Second, although DCSMDA does not rely on any known experimentally verified miRNA-disease relationships, the performance of DCSMDA was not very satisfactory compared with that of several existing methods, such as LRSMDA and WBSMDA [27, 28]. Introducing more reliable measures for the calculations of the disease similarity, miRNA similarity, and lncRNA similarity and developing a more reliable similarity integration method could improve the performance of DCSMDA. Finally, DCSMDA cannot be applied to unknown diseases or miRNAs that are not present in the disease-miRNA or lncRNA-miRNA databases; such genes are

Zhao *et al. BMC Bioinformatics* (2018) 19:141

Page 7 of 14

poorly investigated and have no known disease-lncRNA and lncRNA-miRNA associations. The performance of DCSMDA will be further improved once more known associations are obtained.

Conclusion

In this article, we mainly achieved the following contributions: (1) we constructed a miRNA-lncRNA-disease interactive network based on common assumptions that similar diseases tend to show similar interaction and non-interaction patterns with lncRNAs, and similar miRNAs tend to show similar interaction and non-interaction patterns with lncRNAs; (2) the concept of a distance correlation set was introduced; (3) the sematic disease similarity, functionally similarity (including disease functionally similarity and miRNA functionally similarity) and Gaussian interaction profile kernel similarity (including disease Gaussian interaction profile kernel similarity, miRNA Gaussian interaction profile kernel similarity and lncRNA Gaussian interaction profile kernel similarity) were integrated; (4) the concept of an optimized matrix was introduced by integrating the Gaussian interaction profile kernel similarity of the miRNA pairs and disease pairs; (5) negative samples are not required in DCSMDA; and (6) DCSMDA can be applied to human diseases without relying on any known miRNA-disease associations.

Methods

Known disease-lncRNA associations

Because the number of lncRNA-disease associations is limited and many heterogeneous biological datasets have been constructed, we collected 8842 known disease-lncRNA associations from the MNDR dataset (http://www.bioinformatics.ac.cn/mndr/index.html) and 2934 known disease-lncRNA associations from the LncRNADisease dataset (http://www.cuilab.cn/lncrnadisease). Since the disease names in the LncRNADisease database differ from those in the MNDR dataset, we mapped the diseases in these two disease-lncRNA association datasets to their MeSH descriptors. After eliminating diseases without any MeSH descriptors, merging the diseases with the same MeSH descriptors and removing the lncRNAs that were not present in the lncRNA-miRNA dataset (DS_4) used in this paper, 583 known lncRNA-disease associations (DS_1) were obtained from the LncRNADisease dataset (see Additional file 1), and 702 known lncRNA-disease associations (DS_2) were obtained from the MNDR dataset (see Additional file 2). Furthermore, after integrating the DS_1 and DS_2 datasets and removing the duplicate associations, we obtained the DS_3 dataset, which included 1073 disease-lncRNA associations (see Additional file 3).

Known lncRNA-miRNA associations

To construct the lncRNA-miRNA network, the lncRNA-miRNA association dataset DS_4 was obtained from the starBasev2.0 database (http://starbase.sysu.edu.cn/) in February 2, 2017 and provided the most comprehensive experimentally confirmed lncRNA-miRNA interactions based on large-scale CLIP-Seq data. After the data pre-processing (including the elimination of duplicate values, erroneous data, and disorganized data), removing the lncRNAs that did not exist in the DS_3 dataset and merging the miRNA copies that produced the same mature miRNA, we finally obtained 1883 lncRNA-miRNA associations (DS_4) (see Additional file 4).

Known disease-miRNA associations

To validate the performance of DCSMDA, the known human miRNA-disease associations were downloaded from the latest version of the HMDD database, which is considered the golden-standard dataset. In this dataset, after eliminating the duplicate associations and miRNA-disease associations involved with other diseases or lncRNAs not contained in the DS_3 or DS_4, we finally obtained 3252 high-quality lncRNA-disease associations (DS_5) (see Additional file 5).

Construction of the disease-lncRNA-miRNA interaction network

To clearly demonstrate the process of constructing the disease-lncRNA-miRNA interaction network, we use the disease-lncRNA dataset DS_3 and the lncRNA-miRNA dataset DS_4 as examples. We defined L to represent all the different lncRNA terms in DS_3 and DS_4 and then constructed the disease-lncRNA-miRNA interactive network based on DS_3 and DS_4 according to the following 3 steps:

Step 1 (Construction of the disease-lncRNA network): Let D and L be the number of different diseases and lncRNAs obtained from DS_3, respectively. $S_D = \{d_1, d_2..., d_D\}$ represents the set of all D different diseases in DS_3. $S_L = \{l_1, l_2..., l_L\}$ represents the set of all L different lncRNAs in DS_3, and for any given $d_i \in S_D$ and $l_j \in S_L$, we can construct the $D*L$ dimensional matrix KAM1 as follows:

$$KAM1(i,j) = \begin{cases} 1 & \text{if } d_i \text{ is related to } l_j \text{ in } DS_3 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

Step 2 (Construction of the lncRNA-miRNA network): Let M be the number of different miRNAs obtained from DS_4. $S_M = \{m_1, m_2..., m_M\}$ represents the set of all M different miRNAs in DS_4, and for any given $m_i \in S_M$ and $l_j \in S_L$, we can construct the $M*L$ dimensional matrix KAM2 as follows:

Zhao et al. BMC Bioinformatics (2018) 19:141

Page 8 of 14

$$KAM2(i, j) = \begin{cases} 1 & \text{if } m_i \text{ is related to } l_j \text{ in } DS_4 \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

Step 3 (Constriction of the disease-lncRNA-miRNA interactive network): Based on the disease-lncRNA network and lncRNA-miRNA network, we can obtain the undirected graph $G_3 = (V_3, E_3)$, where $V_3 = S_D \cup S_L \cup S_M = \{d_1, d_2, ..., d_D, l_{D+1}, l_{D+2}..., l_{D+L}, m_{D+L+1}, m_{D+L+2}..., m_{D+L+M}\}$ is the set of vertices, E_3 is the edge set of G_3, and $d_i \in S_D$, $l_j \in S_L$, $m_k \in S_M$. Here, an edge exists between d_i and l_j in $E_3 KAM1(d_i, l_j) = 1$, an edge exists between l_j and m_k in E_3 if $KAM2(m_k, l_j) = 1$. Then, for any given $a, b \in V_3$, we can define the Strong Correlation (SC) between a and b as follows:

$$SC(a, b) = \begin{cases} 1 & \text{if there is an edge between } a \text{ and } b \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

Notably, although we did not use any known disease-miRNA associations, the diseases and miRNAs can still be indirectly linked by integrating the edges between the disease nodes, the lncRNA nodes and edges between the miRNA nodes and lncRNA nodes in G_3.

Disease semantic similarity

We downloaded the MeSH descriptors of the diseases from the National Library of Medicine (http://www.nlm.nih.gov/), which introduced the concept of Categories and Subcategories and provided a strict system for disease classification. The topology of each disease was visualized as a Directed Acyclic Graph (DAG) in which the nodes represented the disease MeSH descriptors, and all MeSH descriptors in the DAG were linked from more general terms (parent nodes) to more specific terms (child nodes) by a direct edge (see Fig. 4). Let $DAG(A) = (A, T(A), E(A))$,
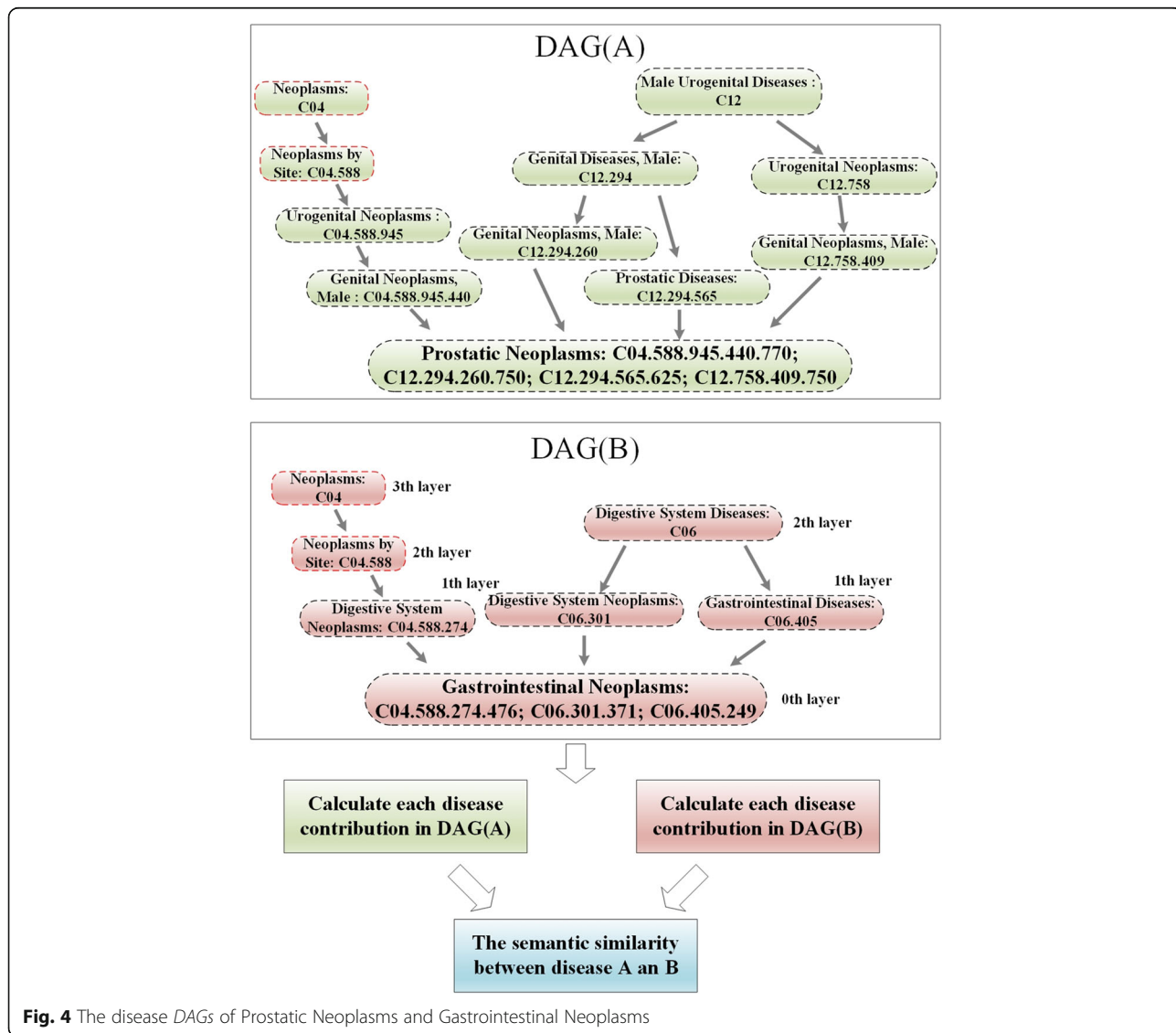
Fig. 4 The disease *DAGs* of Prostatic Neoplasms and Gastrointestinal Neoplasms

Zhao et al. BMC Bioinformatics (2018) 19:141

Page 9 of 14

where A represents disease A, $T(A)$ represents the node set, including node A and its ancestor nodes, and $E(A)$ represents the corresponding edge set. Then, we defined the contribution of disease term d in $DAG(A)$ to the semantic value of disease A as follows:

$$\begin{cases} D_A(d) = 1 & \text{if } d = A \\ D_A(d) = \max\{0.5 * D_A(d^*) | d^* \in children\ of\ d\} & \text{if } d \neq A \end{cases}$$

(4)

For example, the semantic value of the disease 'Gastrointestinal Neoplasms' shown in Fig. 4 is calculated by summing the weighted contribution of 'Neoplasms' (0.125), 'Neoplasms by Site' (0.25), 'Digestive System Diseases' (0.25), 'Digestive System Neoplasms' (0.5), 'Digestive System Neoplasms' (0.5) and 'Gastrointestinal Diseases' (0.5) to 'Gastrointestinal Neoplasms' and the contribution to 'Gastrointestinal Neoplasms' (1) by 'Gastrointestinal Neoplasms'.

Then, the sematic value of disease A can be obtained by summing the contribution from all disease terms in $= DAG(A)$, and the semantic similarity between the two diseases d_i and d_j can be calculated as follows:

$$SSD(d_i, d_j) = \frac{\sum_{d \in (T(d_i) \cap T(d_j))} (D_{d_i}(d) + D_{d_j}(d))}{\sum_{d \in T(d_i)} D_{d_i}(d) + \sum_{d \in T(d_j)} D_{d_j}(d)}$$

(5)

where SSD is the disease semantic similarity matrix.

MiRNA Gaussian interaction profile kernel similarity

Based on the assumption that similar miRNAs tend to show similar interaction and non-interaction patterns with lncRNAs, in this section, we introduce the Gaussian interaction profile kernel used to calculate the network topologic similarity between miRNAs and used the vector $MLP(m_i)$ to denote the ith row of the adjacency matrix $KAM2$. Then, the Gaussian interaction profile kernel similarity for all investigated miRNAs can be calculated as follows:

$$MGS(m_i, m_j) = \exp\left(-\frac{M * \|MLP(m_i) - MLP(m_j)\|^2}{\sum_{i=1}^{M} \|MLP(m_i)\|^2}\right)$$

(6)

where parameter M is the number of miRNAs in DS_4.

Disease Gaussian interaction profile kernel similarity

Based on the assumption that similar diseases tend to show similar interaction and non-interaction patterns with lncRNAs, the Gaussian interaction profile kernel similarity for all investigated diseases can be calculated as follows:

$$DGS(d_i, d_j) = \exp\left(-\frac{D * \|DLP(d_i) - DLP(d_j)\|^2}{\sum_{i=1}^{D} \|DLP(d_i)\|^2}\right)$$

(7)

where parameter D is the number of diseases in DS_3, and $DLP(d_i)$ represent the ith row of the matrix $KAM1$. Then, based on previous work [46], we can improve the predictive accuracy problems by logistic function transformation as follows:

$$FDGS(d_i, d_j) = \frac{1}{1 + e^{-15 * DGS(d_i, d_j) + \log(9999)}}$$

(8)

lncRNA Gaussian interaction profile kernel similarity

Based on the assumption that similar lncRNAs tend to show similar interaction and non-interaction patterns with miRNAs and similar lncRNAs tend to show similar interaction and non-interaction patterns with diseases, the Gaussian interaction profile kernel similarity matrix for all investigated lncRNAs in DS_3 can be computed in a similar way as that for disease, as follows:

$$LGS1(l_i, l_j) = \exp\left(-\frac{L * \|LDP(l_i) - LDP(l_j)\|^2}{\sum_{i=1}^{L} \|LDP(l_i)\|^2}\right)$$

(9)

where parameter L is the number of lncRNAs in DS_3, and $LDP(l_i)$ represents the ith column of the matrix $KAM1$.

Obviously, the Gaussian interaction profile kernel similarity for all investigated lncRNAs in DS_4 can be computed as follows:

$$LGS2(d_i, d_j) = \exp\left(-\frac{L * \|LMP(l_i) - LMP(l_j)\|^2}{\sum_{i=1}^{L} \|LMP(l_i)\|^2}\right)$$

(10)

where $LMP(l_i)$ represents the ith column of the matrix $KAM2$.

Disease functional similarity based on the lncRNAs

To calculate the functional similarity of the diseases, we first constructed the undirected graph $G_1 = (V_1, E_1)$ based on $KAM1$, where $V_1 = S_D \cup S_M = \{d_1, d_2, ..., d_D, l_{D+1}, l_{D+2}, ..., l_{D+M}\}$ is the set of vertices, E_1 is the set of edges, and for any two nodes $a, b \in V_1$, an edge exists between a and b in E_1 if $KAM1(a, b) = 1$. Therefore, we can calculate the similarities between two disease nodes by comparing and integrating the similarities of the lncRNA nodes associated with these two disease nodes based on the assumption that similar diseases tend to

Zhao *et al. BMC Bioinformatics* (2018) 19:141
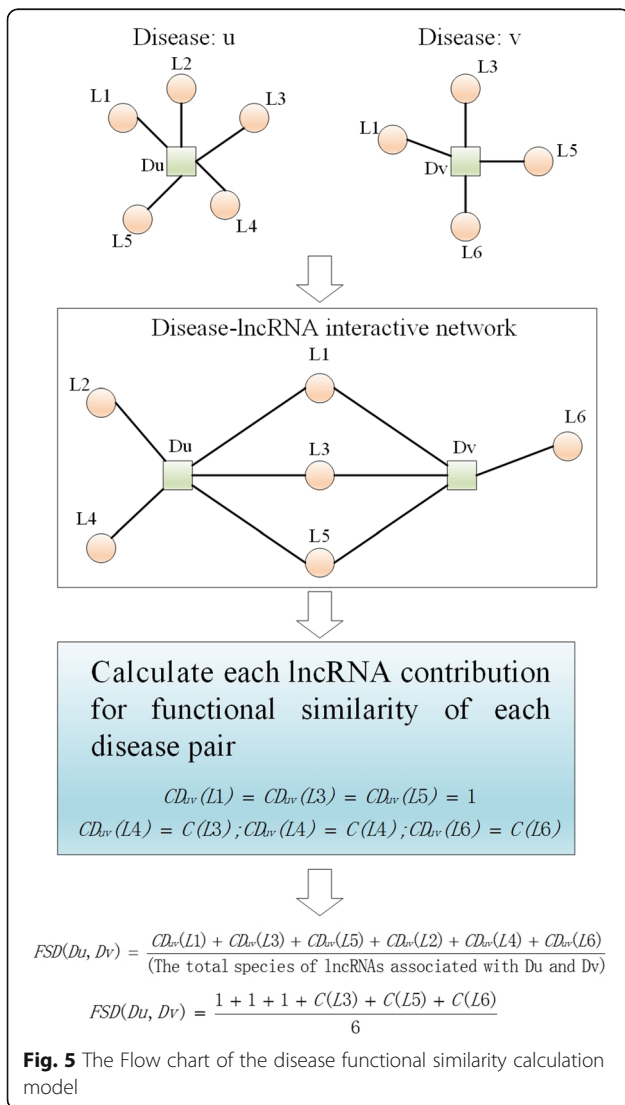
Page 10 of 14

show similar interaction and non-interaction patterns with lncRNAs. The procedure used to calculate the disease functional similarity is shown in Fig. 5.

Because different lncRNA terms in DS_3 may relate to several diseases, assigning the same contribution value to all miRNAs is not suitable, and therefore, we defined the contribution value of each lncRNA as follows:

$$C(l_i) = \frac{\text{The number of } l_i\text{-related edges in } E_1}{\text{The number of all edges in } E_1} \quad (11)$$

Based on the definition of $C(l_i)$, we can define the contribution value of each lncRNA to the functional similarity of each disease pair as follows:

$$CD_{ij}(l_k) = \begin{cases} 1 & \text{if lncRNA } l_k \text{ related to } d_i \text{ and } d_j \text{ simultaneously} \\ C(l_k) & \text{if lncRNA } l_k \text{ only related to } d_i \text{ or } d_j \end{cases} \quad (12)$$

Fig. 5 The Flow chart of the disease functional similarity calculation model

Finally, we can define the functional similarity between diseases d_i and d_j by integrating lncRNAs related to d_i, d_j or both as follows:

$$FSD(d_i, d_j) = \frac{\sum_{l_k \in (D(d_i) \cup D(d_j))} CD_{ij}(l_k)}{|D(d_i)| + |D(d_j)| - |D(d_i) \cap D(d_j)|} \quad (13)$$

where $D(d_i)$ and $D(d_j)$ represent all lncRNAs related to di and d_j in E_1, respectively.

MiRNA functional similarity based on lncRNAs
Based on the assumption that similar miRNAs tend to show similar interaction and non-interaction patterns with lncRNAs, we can also calculate the miRNA functional similarity in the lncRNA-miRNA interactive network. Similar to the procedure used to calculate the disease functional similarity, first, we constructed the undirected graph $G_2 = (V_2, E_2)$, where $V_2 = S_M \cup S_L$ $= \{m_1, m_2,..., l_{M+1}, l_{M+2},..., l_{M+L}\}$ is the set of vertices, E_2 is the set of edges, and for any two nodes a, $b \in V_2$, an edge exists between a and b in E_2 if $KAM2(a, b) = 1$. Then, we defined the contribution of each lncRNA to the functional similarity of each miRNA pair as follows:

$$CM_{ij}(l_k) = \begin{cases} 1 & \text{if lncRNA } l_k \text{ related } m_i \text{ and } m_j \text{ simultaneously} \\ C(l_k) & \text{if lncRNA } l_k \text{ only related } m_i \text{ or } m_j \end{cases} \quad (14)$$

Additionally, we can define the functional similarity between m_i and m_j as follows:

$$FSM(m_i, m_j) = \frac{\sum_{l_k \in (D(m_i) \cup D(m_j))} CM_{ij}(m_k)}{|D(m_i)| + |D(m_j)| - |D(m_i) \cap D(m_j)|} \quad (15)$$

where $D(m_i)$ represents all lncRNAs related to m_i, and $D(m_j)$ represents lncRNAs relate to m_j in E_2.

Integrated similarity
The processes used to calculate the integrated similarities of the diseases, lncRNAs and miRNAs are illustrated in Fig. 6. Combining the disease semantic similarity, the disease Gaussian interaction profile kernel similarity and the disease functional similarity mentioned above, we can construct the disease integrated similarity matrix FDD as follows:

$$FDD = \frac{SSD + FDGS + FSD}{3} \quad (16)$$

Additionally, based on the miRNA Gaussian interaction profile kernel similarity and the miRNA functional similarity, we can construct the miRNA integrated similarity matrix FMM as follows:

Zhao *et al. BMC Bioinformatics* (2018) 19:141

Page 11 of 14
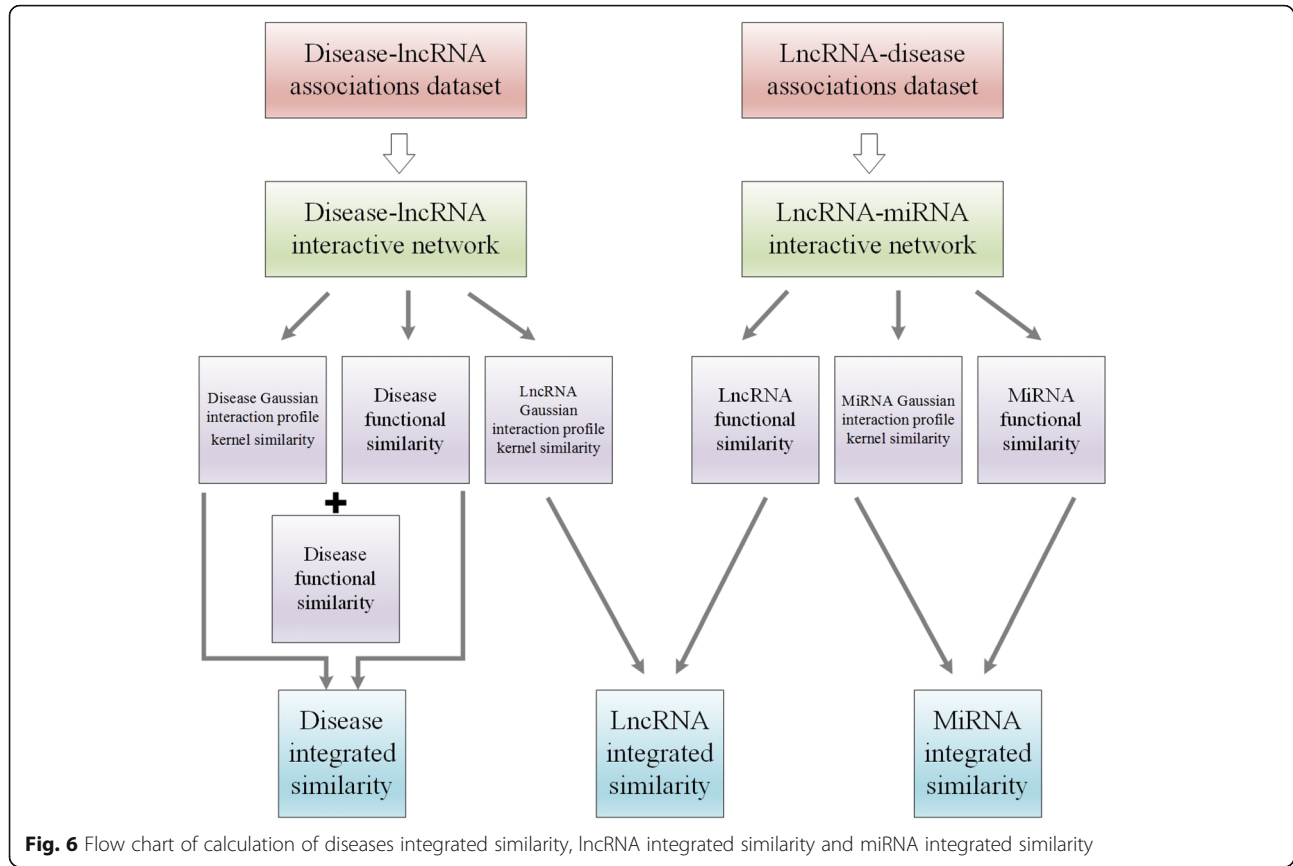
Fig. 6 Flow chart of calculation of diseases integrated similarity, lncRNA integrated similarity and miRNA integrated similarity

$$FMM = \frac{MGS + FSM}{2} \tag{17}$$

Furthermore, based on the Gaussian interaction profile kernel similarity matrices *LGS1* and *LGS2*, we can construct the lncRNA integrated similarity matrix *FLL* as follows:

$$FLL = \frac{LGS1 + LGS2}{2} \tag{18}$$

Prediction of disease-miRNA associations based on a distance correlation set

In this section, we developed a novel computational method, i.e., DCSMDA, to predict potential disease-miRNA associations by introducing a distance correlation set based on the following assumptions: similar diseases tend to show similar interaction and non-interaction patterns with lncRNAs, and similar lncRNAs tend to show similar interaction and non-interaction patterns with miRNAs. As illustrated in Fig. 7, the DCSMDA procedure consists of the following 5 major steps:
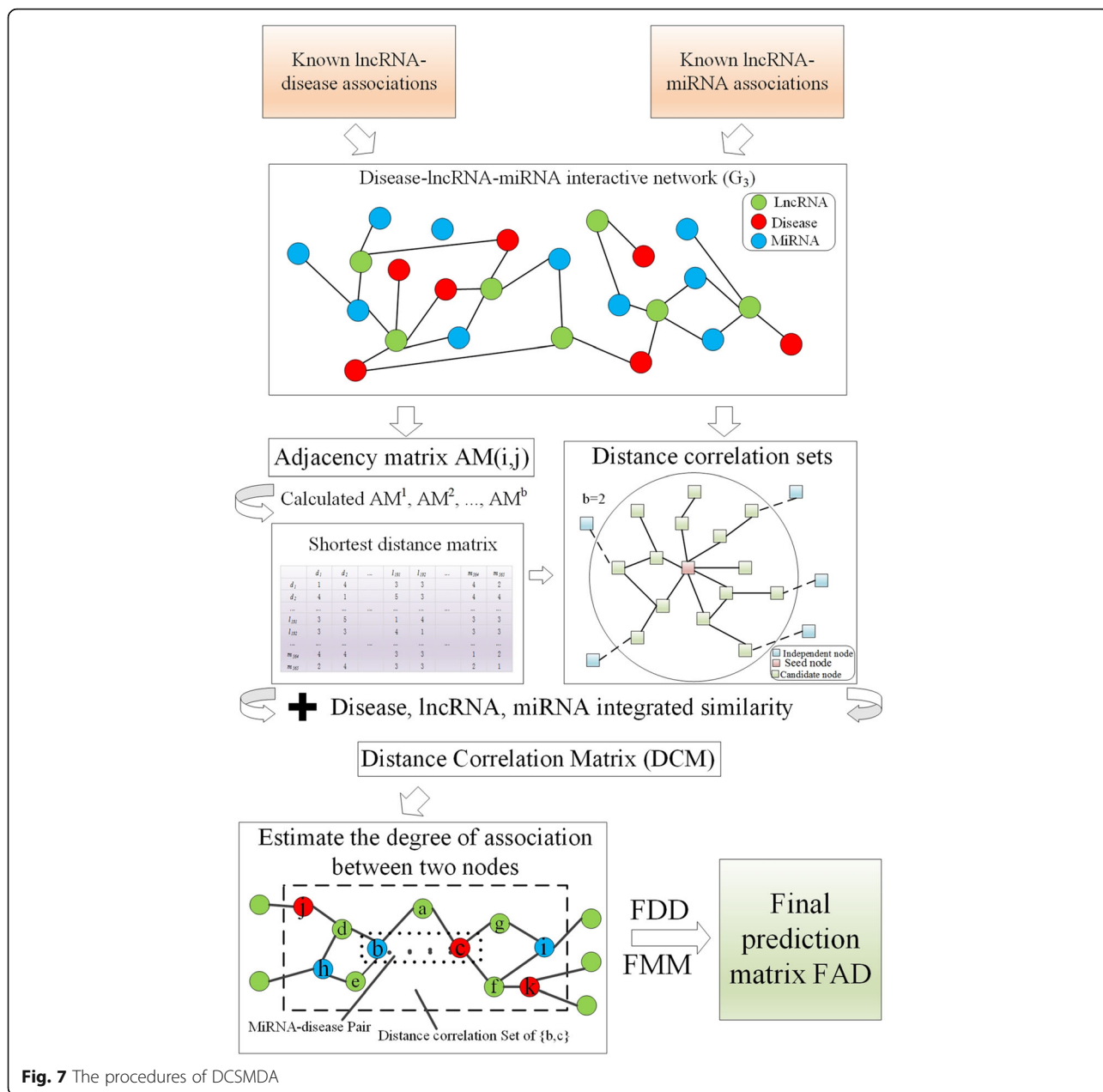
Step 1 (Construction of the adjacency matrix based on G_3): First, we construct a $(D + L + M) * (D + L + M)$ Adjacency Matrix (*AM*) based on the undirected graph G_3 and *SC*, and then for any two nodes v_i, $v_j \in V_3$, we define the $AM(i, j)$ as follows:

$$AM(i,j) = \begin{cases} SC(d_i, d_j), & if \ i \in [1,D] \text{ and } j \in [1,D]. \\ SC(d_i, l_j), & if \ i \in [1,D] \text{ and } j \in [D,D+L]. \\ SC(d_i, m_j), & if \ i \in [1,D] \text{ and } j \in [D+L,D+L+M]. \\ SC(m_i, d_j), & if \ i \in [D,D+L] \text{ and } j \in [1,D]. \\ SC(m_i, m_j), & if \ i \in [D,D+L] \text{ and } j \in [D,D+L]. \\ SC(l_i, m_j), & if \ i \in [D,D+L] \text{ and } j \in [D+L,D+L+M]. \\ SC(l_i, d_j), & if \ i \in [D+L,D+L+M] \text{ and } j \in [1,D]. \\ SC(l_i, m_j), & if \ i \in [D+L,D+L+M] \text{ and } j \in [D,D+L]. \\ SC(l_i, m_j), & if \ i \in [D+L,D+L+M] \text{ and } j \in [D+L,D \\ & \qquad +L+M] \end{cases} \tag{19}$$

where $i \in [1, \ D+L+M]$ and $j \in [1, \ D+L+M]$, and to calculate the shortest distance matrix in step 2, we define $AM \ (i, j) = 1$ if $i = j$.

Step 2 (Construction of the shortest distance matrix based on adjacency matrix *AM*): First, we set parameter b to control the bandwidth of the distance correlation set and let b be a pre-determined positive integer, and then, we can obtain b matrices, such as AM^1, AM^2,..., AM^b, based on the above formula (19), and the Shortest Path Matrix is calculated as follows:

$$SPM(i,j) = \begin{cases} 1, & if \ AM(i,j) = 1 \\ k, & otherwise \end{cases} \tag{20}$$

where $i \in [1, \ D + M + L]$, $j \in [1, \ D + M + L]$, $k \in [2, \ b]$, and k satisfies the following: $AM^{\ k}(i, j) \neq 0$, while $AM^{\ 1}(i, j) = AM^{\ 2}(i, j) = ... = AM^{\ k-1}(i, j) = 0$.

Zhao *et al. BMC Bioinformatics* (2018) 19:141

Page 12 of 14

Fig. 7 The procedures of DCSMDA

Step 3 (Calculation of distance correlation sets and distance coefficient of each node pair in G_3):

For each node $v_i \in V_3$, we can obtain distance correlation set $DCS(i)$ according to the shortest distance matrix as follows:

$$DCS(i) = \{v_j | r \geq SPM(i,j) > 0\} \quad (21)$$

where $DCS(i)$ of each node contains itself and all nodes with the shortest distance less than b.

For instance, in the disease-miRNA-lncRNA interaction network illustrated in Fig. 7, DCS (seed node) is all candidate nodes when b is set to 2.

Then, we can calculate the distance coefficient (DC) of the node pair (v_i, v_j) as follows:

$$P(i,j) = \begin{cases} SPM(i,j)^{b+1}, if \ i \in DCS(j) \ or \ j \in DCS(i) \\ 0, \quad\quad\quad otherwise \end{cases}$$

$$(22)$$

Furthermore, we can construct a Distance Correlation Matrix (DCM) based on the disease integrated similarity, the lncRNA integrated similarity, and the miRNA integrated similarity as follows:

Zhao *et al. BMC Bioinformatics* (2018) 19:141

Page 13 of 14

$$DCM(i,j) = \begin{cases} P(i,j) * \exp(FDD(i,j)), & \text{if } i \in [1,D] \text{ and } j \in [1,D]. \\ P(i,j) * \exp(FLL(i,j)), & \text{if } i \in [D, D+L] \text{ and } j \in [D, D+L]. \\ P(i,j) * \exp(FMM(i,j)), & \text{if } i \in [D+L, D+L+M] \text{ and } j \in [D+L, D+L+M] \\ P(i,j) * \dfrac{SPM(i,j)}{b}, & \text{otherwise} \end{cases} \quad (23)$$

where $i \in [1, D+L+M]$ and $j \in [1, D+L+M]$.

Step 4 (Estimation of the association degree between a pair of nodes): Based on formula (23), we can estimate the association degree between v_i and v_j as follows:

$$PM(i,j) = \frac{\sum_{k=1}^{D+L+M} DCM(i,k) + \sum_{k=1}^{D+L+M} DCM(k,j)}{D+L+M} \quad (24)$$

Thus, we can obtain prediction matrix *PM*, where the entity *PM (i, j)* in row *i* column *j* represents the predicted association between node v_i and v_j.

Step 5 (Calculation of the final prediction result matrix between the miRNAs and diseases): Let $PM = \begin{bmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{bmatrix}$, where C_{11} is a *D×D* matrix, C_{12} is a *D×L* matrix, C_{13} is a *D×M* matrix, C_{21} is an *L×D* matrix, C_{22} is an *L×L* matrix, C_{23} is an L×M matrix, C_{31} is an M×D matrix, C_{32} is an *M×L* matrix and C_{33} is an *M×M* matrix. Obviously, C_{13} is our predicted result, which provides the association probability between each disease and miRNA. A previous study [27] demonstrated that the Gaussian interaction profile kernel similarity is a high-efficiency tool for optimizing the result of prediction, and therefore, we used the miRNA Gaussian interaction profile kernel similarity and the disease Gaussian interaction profile kernel similarity to optimize the result of the DCSMDA as follows:

$$FAD = FDD * C_{13} * FMM \quad (25)$$

where the matrix FAD denotes the relationship between the miRNA-disease pairs.

Additional files

Additional file 1: The known lncRNA-disease associations for constructing the DS_1. We list 583 known lncRNA-disease associations which were collected from LncRNAdisease dataset to construct the DS_1. (XLS 58 kb)

Additional file 2: The known lncRNA-disease associations for constructing the DS_2. We list 702 known lncRNA-disease associations which were collected from MNDR dataset to construct the DS_2. (XLS 63 kb)

Additional file 3: The integrated lncRNA-disease associations for constructing the DS_3. We list 1073 lncRNA-disease associations which were collected by integrating the datasets of DS_1 and DS_2. (XLS 83 kb)

Additional file 4: The known lncRNA-miRNA associations for constructing the DS_4. We list 1883 known lncRNA-miRNA associations which were collected from starBasev2.0 database to construct the DS_4. (XLS 123 kb)

Additional file 5: The known miRNA-disease associations for constructing the DS_5. We list 3252 high-quality miRNA-disease associations which were collected from HMDD database to validate the performance of our method. (XLS 191 kb)

Abbreviations
AUC: Areas under ROC curve; DCSMDA: Distance Correlation Set is developed to predict MiRNA-Disease Associations; FPR: False positive rates; GO: Gene Ontology; miRNA: MicroRNA; ncRNA: Noncoding RNA; ROC: Receiver-operating characteristics; TPR: True positive rates; LOOCV: Leave-One Out Cross Validation

Availability of data and materials
All data generated or analyzed during this study are included in this published article [Additional file 1, Additional file 2, Additional file 3, Additional file 4 and Additional file 5].

Author's contributions
HCZ conceived the study. HCZ, LAK and LW developed the method. PYP and ZWX implemented the algorithms. HCZ and TRP analyzed the data. LW supervised the study. HCZ and LW wrote the manuscript. ZLW, PYP and LW reviewed and improved the manuscript, ZLW provided supplementary data. All authors read and approved the final manuscript.

Ethics approval and consent to participate
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
[1]College of Computer Engineering & Applied Mathematics, Changsha University, Changsha 410001, Hunan, People's Republic of China. [2]Key Laboratory of Intelligent Computing & Information Processing (Xiangtan University), Ministry of Education, China, Xiangtan 411105, Hunan, People's Republic of China. [3]Department of Computer Science, Lakehead University, Thunder Bay, ON P7B5E1, Canada. [4]Department of Computer Science, Princeton University, Princeton, New Jersey, USA. [5]College of Information

Zhao *et al. BMC Bioinformatics* (2018) 19:141

Page 14 of 14

Engineering, Xiangtan University, Xiangtan, Hunan, People's Republic of China.

References

1. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009;10(1):57–63.
2. Crick FHC, Barnett L, Brenner S, Watts-Tobin RJ. General nature of the genetic code for proteins. Nature. 1961;192(4809):1227–32.
3. Mattick JS, Makunin IV. Non-coding RNA. Hum Mol Genet. 2006;15(suppl_1):R17.
4. Esteller M. Non-coding RNAs in human disease. Nat Rev Genet. 2011;12(12):861–74.
5. Mattick JS, Rinn JL. Discovery and annotation of long noncoding RNAs. Nat Struct Mol Biol. 2015;22(1):5.
6. Ambros V. The functions of animal micrornas. Nature. 2004;431(7006):350.
7. Cheng AM, Byrom MW, Shelton J, Ford LP. Antisense inhibition of human mirnas and indications for an involvement of mirna in cell growth and apoptosis. Nucleic Acids Res. 2005;33(4):1290–7.
8. Taguchi Y. Inference of target gene regulation via mirnas during cell senescence by using the mirage server. Aging Dis. 2012;3(4):301.
9. Peng H, Lan C, Zheng Y, Hutvagner G, Tao D, Li J. Cross disease analysis of co-functional microrna pairs on a reconstructed network of disease-gene-microrna tripartite. Bmc Bioinformatics. 2017;18(1):193.
10. Weber MJ. New human and mouse microrna genes found by homology search. FEBS J. 2005;272(1):59.
11. Thum T, Gross C, Fiedler J, Fischer T, Kissler S, Bussen M, et al. Microrna-21 contributes to myocardial disease by stimulating map kinase signalling in fibroblasts. Nature. 2008;456(7224):980–4.
12. Cogswell JP, Ward J, Taylor IA, Waters M, Shi Y, Cannon B, et al. Identification of mirna changes in alzheimer's disease brain and csf yields putative biomarkers and insights into disease pathways. J Alzheimers Dis. 2008;14(1):27–41.
13. Corsten MF, Dennert R, Jochems S, Kuznetsova T, Devaux Y, Hofstra L, et al. Circulating microrna-208b and microrna-499 reflect myocardial damage in cardiovascular disease. Circ Cardiovasc Genet. 2010;3(6):499.
14. Ikeda S, Kong SW, Lu J, Bisping E, Zhang H, Allen PD, et al. Altered microrna expression in human heart disease. Physiol Genomics. 2007;31(3):367–73.
15. Lu M, Zhang Q, Deng M, Miao J, Guo Y, Gao W, et al. An analysis of human microrna and disease associations. PLoS One. 2008;3(10):e3420.
16. Chen X, Liu MX, Yan GY. Rwrmda: predicting novel human microrna-disease associations. Mol BioSyst. 2012;8(10):2792.
17. Li Y, Qiu C, Tu J, Geng B, Yang J, Jiang T, et al. HMDD v2.0: a database for experimentally supported human microrna and disease associations. Nucleic Acids Res. 2014;42(Database issue):1070–4.
18. 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.
19. Jiang Q, Wang Y, Hao Y, Juan L, Teng M, Zhang X, et al. Mir2disease: a manually curated database for microrna deregulation in human disease. Nucleic Acids Res. 2009;37(1):D98–104.
20. Zou Q, Li J, Hong Q, Lin Z, Wu Y, Shi H, et al. Prediction of microrna-disease associations based on social network analysis methods. Biomed Res Int. 2015;2015(10):810514.
21. You ZH, Wang LP, Chen X, et al. PRMDA: personalized recommendation-based MiRNA-disease association prediction[J]. Oncotarget. 2017;8(49):85568-83.
22. Shi H, Xu J, Zhang G, Xu L, Li C, Wang L, et al. 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):1–12.
23. Jiang Q, Hao Y, Wang G, Juan L, Zhang T, Teng M, et al. Prioritization of disease micrornas through a human phenome-micrornaome network. BMC Syst Biol. 2010;4(S1):S2.
24. Han K, Xuan P, Ding J, Zhao ZJ, Hui L, Zhong YL. Prediction of disease-related micrornas by incorporating functional similarity and common association information. Gen Mol Res Gmr. 2014;13(1):2009–19.
25. Xuan P, Han K, Guo M, Guo Y, Li J, Ding J, et al. Prediction of micrornas associated with human diseases based on weighted k most similar neighbors. PLoS One. 2013;8(9):e70204.
26. Xuan P, Han K, Guo Y, Li J, Li X, Zhong Y, et al. Prediction of potential disease-associated micrornas based on random walk. Bioinformatics. 2015;31(11):1805–15.
27. Chen X, Yan GY. Semi-supervised learning for potential human microrna-disease associations inference. Sci Rep. 2014;4:5501.
28. Chen X, Yan CC, Zhang X, You ZH, Deng L, Liu Y, et al. Wbsmda: within and between score for mirna-disease association prediction. Sci Rep. 2016;6:21106.
29. Wang Y, Chen L, Chen B, Li X, Kang J, Fan K, et al. Mammalian ncrna-disease repository: a global view of ncrna-mediated disease network. Cell Death Dis. 2013;4(8):e765.
30. Chen G, Wang Z, Wang D, Qiu C, Liu M, Chen X, et al. lncrna-disease: a database for long-non-coding rna-associated diseases. Nucleic Acids Res. 2013;41(Database issue):983–6.
31. Chen X. Predicting lncrna-disease associations and constructing lncrna functional similarity network based on the information of mirna. Sci Rep. 2015;5:13186.
32. Huang WT, Guo XQ, Dai JP, Chen RS. Microrna and lncrna in neurodegenerative diseases*: microrna and lncrna in neurodegenerative diseases. Prog Biochem Biophys. 2010;37(8):826–33.
33. Guo L, Peng Y, Meng Y, et al. Expression profiles analysis reveals an integrated miRNA-lncRNA signature to predict survival in ovarian cancer patients with wild-type BRCA1/2. Oncotarget. 2017;8(40):68483.
34. Spiess PE, Dhillon J, Baumgarten AS, Johnstone PA, Giuliano AR. Pathophysiological basis of human papillomavirus in penile cancer: key to prevention and delivery of more effective therapies. CA-A Cancer J Clinicians. 2016;6:481–95.
35. Ruprecht B, Zaal EA, Zecha J, Wu W, Berkers CR, Kuster B, Lemeer S. Lapatinib resistance in breast Cancer cells is accompanied by phosphorylation-mediated reprogramming of glycolysis. Cancer Res. 2017;77(8):1842–53.
36. Barton MK. Local consolidative therapy may be beneficial in patients with oligometastatic non-small cell lung cancer. CA-A Cancer J Clinicians. 2017;2:89–90.
37. Jiang J, Jia P, Zhao Z, Shen B. Key regulators in prostate cancer identified by co-expression module analysis. BMC Genomics. 2014;15(1):1015.
38. Medina-Villaamil V, Martínez-Breijo S, Portela-Pereira P, Quindós-Varela M, Santamarina-Caínzos I, Antón-Aparicio LM, et al. Circulating micrornas in blood of patients with prostate cancer. Actas Urol Esp. 2014;38(10):633–9.
39. Cai C, Chen QB, Han ZD, Zhang YQ, He HC, Chen JH, et al. Mir-195 inhibits tumor progression by targeting rps6kb1 in human prostate cancer. Clin Cancer Res. 2015;21(21):4922.
40. Bozok ÇV, Tetik VA, Düzgün Z, Tezcanlı KB, Açıkgöz E, Aktuğ H, et al. Mir-15a enhances the anticancer effects of cisplatin in the resistant non-small cell lung cancer cells. Tumor Biol. 2016;37(2):1739–51.
41. Liu B, Qu J, Xu F, Guo Y, Wang Y, Yu H, et al. Mir-195 suppresses non-small cell lung cancer by targeting chek1. Oncotarget. 2015;6(11):9445–56.
42. Li H, Lan H, Zhang M, An N, Yu R, He Y, et al. Effects of mir-424 on proliferation and migration abilities in non-small cell lung cancer a549 cells and its molecular mechanism. Zhongguo Fei Ai Za Zhi. 2016;19:571–6.
43. Sun YP, Lu F, Han XY, et al. MiR-424 and miR-27a increase TRAIL sensitivity of acute myeloid leukemia by targeting PLAG1. Oncotarget. 2016;7(18):25276-90.
44. Cimmino A, Calin GA, Fabbri M, Iorio MV, Ferracin M, Shimizu M, et al. Mir-15 and mir-16 induce apoptosis by targeting bcl2. Proc Natl Acad Sci U S A. 2005;102(39):13944.
45. Zhao TF, Jia HZ, Zhang ZZ, Zhao XS, Zou YF, Zhang W, et al. Lncrna h19 regulates id2 expression through competitive binding to hsa-mir-19a/b in acute myelocytic leukemia. Mol Med Rep. 2017;16(3):3687.
46. Vanunu O, Magger O, Ruppin E, Shlomi T, Sharan R. Associating genes and protein complexes with disease via network propagation. PLoS Comput Biol. 2010;6(1):e1000641.