SCIENTIFIC REPORTS

OPEN

Prediction of potential disease-associated microRNAs by composite network based inference

Bin-Sheng He[1], Jia Qu[2] & Min Chen[3]

MicroRNAs (miRNAs) act a significant role in multiple biological processes and their associations with the development of all kinds of complex diseases are much close. In the research area of biology, medicine, and bioinformatics, prediction of potential miRNA-disease associations (MDAs) on the base of a variety of heterogeneous biological datasets in a short time is an important subject. Therefore, we proposed the model of Composite Network based inference for MiRNA-Disease Association prediction (CNMDA) through applying random walk to a multi-level composite network constructed by heterogeneous dataset of disease, long noncoding RNA (lncRNA) and miRNA. The results showed that CNMDA achieved an AUC of 0.8547 in leave-one-out cross validation and an AUC of 0.8533+/−0.0009 in 5-fold cross validation. In addition, we employed CNMDA to infer novel miRNAs for kidney neoplasms, breast neoplasms and lung neoplasms on the base of HMDD v2.0. Also, we employed the approach for lung neoplasms on the base of HMDD v1.0 and for breast neoplasms that have no known related miRNAs. It was found that CNMDA could be seen as an applicable tool for potential MDAs prediction.

MicroRNAs (miRNAs) is a kind of short noncoding RNA (ncRNA) molecules with about 22 nucleotides in length which can regulate complementary messenger RNAs[1]. Unlike the miRNAs, long noncoding RNAs (lncRNAs) are a sort of heterogeneous ncRNAs with about 200 nucleotides and usually show less sequence conservation. Accumulating evidence indicates that miRNAs are participated in a wide variety of life process of cells, such as proliferation[2], development[3], aging[4], viral infection[5], metabolism[4,6] and so on[5,7]. It is no surprise that miRNAs are closely related to a number of clinically important diseases[8,9]. For example, miR-335 and miR-126 were proved to be metastasis suppressor miRNAs in human breast cancer[10]. In addition, previous study also confirmed that the differential expression of miR-21, -31, -143 and -145 is closely participate in clinic pathologic features of colorectal cancer[11]. Therefore, identification of disease-related miRNAs would be beneficial for disease diagnosis, treatment, and prevention[12]. Currently, unlike traditional time-consuming biological experiments, adopting validation to the predicted miRNA-disease associations (MDAs) obtained from calculation models could reduce a lot of time and cost. Therefore, it is very significant to propose effective calculation models to infer potential MDAs[13–17].

According to the idea that miRNAs with similar functions are usually relevant to similar diseases and the reverse is also true. some researchers built elaborate computational models for the identification of potential MDAs on the basis of known MDAs in databases only. For example, Li *et al.*[18] developed a computational approach based on matrix completion, in which the adjacency matrix constructed from known MDAs was updated to gain final association scores of each miRNA-disease pair. Considering various types of known MDAs, Chen *et al.*[19] constructed an restricted boltzmann machine (RBM) model to further predict four kinds of MDAs.

Based on the information of known MDAs and the corresponding similarity information of diseases and miRNAs, Chen *et al.*[20] developed an effective method via combining all those information to construct a heterogeneous graph and then further inferred MDAs with the consideration of paths between miRNA nodes and disease nodes. Besides, this method could also be implemented to predict for new diseases (miRNAs). Through integrating the distribution information of *k* most similar neighbors per miRNA and the corresponding functional similarity between the miRNA and its neighbors, Xuan *et al.*[21] proposed a reliable computational approach to infer novel MDAs. However, HDMP cannot predict disease-related miRNAs for new diseases. After computing

[1]The First Affiliated Hospital, Changsha Medical University, Changsha, 410219, China. [2]School of Information and Control Engineering, China University of Mining and Technology, Xuzhou, 221116, China. [3]College of Computer Science and Technology, Hunan Institute of Technology, Hengyang, 421002, China. Correspondence and requests for materials should be addressed to J.Q. (email: TB17060015B4@cumt.edu.cn) or M.C. (email: chenmin@hnit.edu.cn)
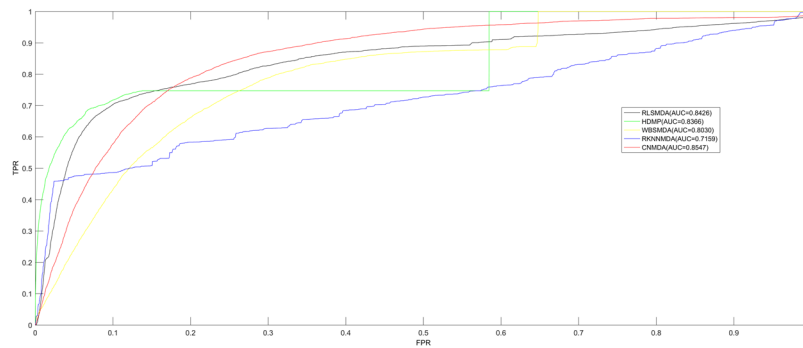
Figure 1. CNMDA got better AUCs of 0.8547 in the LOOCV in comparison of other four calculation approaches (RLSMDA, HDMP, WBSMDA, RKNNMDA).

miRNAs functional similarity (MFS), Xuan *et al.*[22] proposed a prediction model via implementing random walk on constructed miRNA functional similarity network in which they assigned larger transition weights to marked nodes. At last, probability association scores of each disease-miRNA pair would be obtained and ranked. A calculation model was further built by Chen *et al.*[17] in which miRNA's *k*-nearest-neighbors (KNNs) and disease's KNNs were respectively searched and then these KNNs would be ranked according to support vector machine. After that, they finally got all potential MDAs with weighted voting. Under the framework of semi-supervised learning, a novel model[23] was presented for MDAs prediction via combining the optimal solutions in the miRNA space and disease space. Recently, Chen *et al.*[24] proposed another prediction model through calculating within-score and between-score for both miRNAs and diseases which were then combined to obtain the final MDA scores.

Also, researchers put forward some other calculation approaches via considering relevant genes or proteins as a bridge to predict novel MDAs. For example, using a discrete probability distribution of hypergeometric, Jiang *et al.*[25] presented a prediction model on the basis of the constructed integrated network. By connecting miRNAs to diseases with the proteins as a bridge between them, a calculation model was employed by Mork *et al.*[26] through using a scoring scheme, which can greatly increase the model's efficiency. Furthermore, Shi *et al.*[27] implemented random walk on a built protein similarity network to identify MDAs.

By combining the known MDAs network and MFS network, a new calculating method was studied by Chen *et al.*[28] by the analyzed of random walk with restart (RWR). It is worth noting that RWR is a very effective model for MDAs prediction. By adopting RWR, a novel model named Composite Network based inference for MiRNA-Disease Association prediction (CNMDA) was presented in the light of a multi-level network which was built by combination of Gaussian interaction profile kernel similarity (GIPKS) for lncRNAs, integrated similarity for miRNAs (ISMs) and diseases (ISDs), known MDAs, lncRNA-disease associations (LDAs) and miRNA-lncRNA interactions (MLIs). In addition, leave-one-out cross validation (LOOCV) and 5-fold cross validation were adopted in this paper to assess CNMDA's effectiveness. It could be seen that the AUCs of LOOCV and 5-fold cross validation were respectively 0.8547 and 0.8533+/−0.0009. As for case studies, CNMDA was carried out on kidney neoplasms (KN), breast neoplasms (BN) and lung neoplasms (LN) to infer its associated miRNAs based on HMDD v2.0[29]. Also according to HMDD v2.0, we further infer novel miRNAs for BN after hiding its known associated miRNAs. At last, we carried out the case studies based on HMDD v1.0[30] to infer LN-related miRNAs. Based on the above results, the effectiveness of CNMDA for MDAs prediction was validated.

Results

Cross validation. In this paper, we carried out LOOCV and 5-fold cross validation to assess CNMDA's prediction accuracy according to HMDD v2.0[29] and then made comparison between CNMDA and four other classical computational models: RLSMDA[23], HDMP[21], WBSMDA[24] and RKNNMDA[17] (See Fig. 1). In LOOCV, test sample is one of the 5430 MDAs; training samples are the rest of 5429 known MDAs; candidate samples are those unlabeled 184155 miRNA-disease pairs. When each known MDA was taken to be the test sample, we would get association scores for all miRNA-disease pairs after implementing MCMDA and then the ranking of test sample among the candidate samples would be gained based on their association scores. We would say that the model makes a correctly prediction if the test sample ranked higher than the set threshold. Finally, we drew Receiver-Operating Characteristics (ROC) curve through computing the ratio of true positive rate to false positive rate. To evaluate CNMDA's performance, we computed area under the ROC curve (AUC). If AUC = 1, CNMDA would possess perfect performance; If AUC = 0.5, CNMDA could only predict randomly. As a result, CNMDA, RLSMDA, HDMP, WBSMDA, RKNNMDA obtained AUCs of 0.8547 (0.8533+/−0.0009), 0.8426 (0.8569+/−0.0020), 0.8366 (0.8342+/−0.0010), 0.8030 (0.8185+/−0.0009) and 0.7159 (0.6723+/−0.0027) in the LOOCV (5-fold cross validation), respectively. Through comparative analysis with other method, the reliability and effectiveness of CNMDA for identification of potential MDAs were proved.

Case studies. Three different case studies were also implemented to assess CNMDA' performance. In the first case study, CNMDA was employed to predict KN-related miRNAs based on HMDD v2.0. Further, another two reliable MDA databases (dbDEMC and miR2Disease) would be utilized to validate the top 50 identified outcomes. In the second case study, we respectively inferred BN-associated miRNAs and BN-associated miRNAs

after removing all known BN-associated miRNAs in HMDD v2.0. In the third kind of case studies, CNMDA was adopted to predict for LN according to associations in HMDD v1.0 and v2.0, respectively.

KN is a disease caused by cellular metabolic disorders[31]. If kidney tumors are detected and treated early and localized in the kidney, Patients would have a good disease-specific survival rate. Otherwise, patients have only an 18% two-year survival rate when they present with terminal disease[32]. With recent researches and studies, about two hundred and fifty thousand renal tumor patients are newly diagnosed annually, and KN' morbidity and mortality continue to increase[33]. Many miRNAs related to KN have been found based on a large number of biological experiments. For example, in renal cell carcinoma (RCC), up regulation of miR-21 is related to kidney cancer that with lower survival rate[34]. Through targeting MMP-9 in RCC, miRNA-133b can suppress cell proliferation, migration and invasion[35]. Finally, we implemented CNMDA for potential KN-related miRNA prediction. It was found that 8 of the first 10 and 37 of the first 50 miRNAs were verified (See Supplementary Table 1). we also provided the whole scores of potential MDAs on the base of HMDD v2.0 (See Supplementary Table 2).

BN is a major chronic disease affecting adult women and detected breast neoplasms can be removed surgically[36]. However, if people with BN have not been detected, BN may develop into a life-threatening clinical recurrence in the next 5, 10, 15, or more years[37]. Recent experimental studies have provide evidences that miRNA-195 may work as latent biomarker for early BN detection[38]. To find the novel biomarkers for BN for the treatment of the disease is significant. In the second, we employed CNMDA for potential BN-related miRNA prediction. It was found that 5 of the first 10 and 31 of the first 50 miRNAs were verified (See Supplementary Table 3). Also, we implemented CNMDA for the prediction of BN by hiding all its confirmed associations in HMDD v2.0. This means that we would remove all known BN-associated miRNAs and predict potential BN-associated miRNAs based on other known associations and corresponding similarity information. Supplementary Table 4 presents the top 50 predicted outcomes and their verification evidences. As a result, 9 of the first 10 and 41 of the first 50 miRNAs were confirmed (See Supplementary Table 4).

LN is the primary reason of cancer deaths on a global scale[39]. The genetic and epigenetic damage caused by tobacco smoke is the main cause of the disease[40]. Obviously, it is urgent to find a more therapy systemic[39]. In squamous cell carcinoma, miR-126 have been verified to be down regulated and two miRNAs of miR-185∗, miR-125a-5p were up regulated[39]. MiR-205 were expressed differently in the non-small cell lung carcinoma (NSCLC)[40]. In order to test the stability of CNMDA, we employed CNMDA based on the associations in HMDD v2.0 and HMDD v1.0[30], respectively. It was found that 20 and 28 of first 50 associated miRNAs for LN have been verified, respectively (See Supplementary Tables 5 and 6).

As seen in the results above, we can arrival at a conclusion that CNMDA possesses excellent predictive performance for the novel MDAs prediction.

Discussions

As overwhelming evidences expounded that miRNAs are participated in all sorts of diseases. The development of new calculation approaches for predicting MDAs in a short time is important to further experimental validation. Accordingly, it is now possible to confirmed novel MDAs using biological experiments with low time and cost. Existing models are usually proposed based on four different calculation mechanisms[41]. Some scoring functions were constructed to prioritize disease-related miRNAs through carrying out probability distribution. Complex network algorithm-based prediction models were introduced through establishing complex network based on various data that are collected or calculated from different perspectives. Machine learning-based prediction models were introduced by using powerful machine learning algorithms. Moreover, multiple biological information-based models were put forward through constructing intermediate medium associations based on various biological datasets. We put forward the computing method of CNMDA to infer novel MDAs. In the model, we implemented RWR on a multi-level composite network that was built through combining collected and calculated data (ISD, ISM, GIPKS for lncRNAs, experimentally validated MDAs, MLIs and LDAs). From the evaluation results, it can be seen that the accuracy of our prediction model was superior in the comparison with other four models.

The main merits for the effective performance of CNMDA are as follows: Through taking advantage of multi-source information based on reliable database, it is no surprise that the integration strategy of CNMDA could predict potential MDAs effectively. Secondly, in comparison of local network information, RWR is an iterative process based on global network for the MDAs prediction. The attractive properties of global network information have been proved in the identification for potential disease-gene associations, MDAs[41,42], LDAs[43] and drug-target interaction[44]. Furthermore, CNMDA could identify novel diseases that have no known associated miRNAs. At last, the implementation of CNMDA only needs positive samples as training data. Since there is no known negative sample information, the forecasting precision of CNMDA is more convincing. However, some limitations also exist in the computation model of CNMDA. For example, the number of experimentally determined MDAs, LDAs and MLI is insufficient. For the number of known MDAs, only 5430 known MDAs were collected. The more the known MDAs, the higher forecasting precision the model. Importantly, the current forecasting precision still needs to be improved according to the evaluation of LOOCV.

Methods

MiRNA-disease associations.   Experimentally confirmed MDAs used in this paper were come from high-quality database[29]. Through constructing a adjacency matrix W_{dm} to indicate the 5430 known MDAs, we made use of variables nm and nd to express the total amount of miRNAs and diseases in the known MDAs dataset, respectively.

$$W_{dm}(i, j) = \begin{cases} 1, & \text{if } miRNA \ m(j) \text{ is related to disease } d(i) \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

LncRNA-disease associations. The known LDAs was from the LncRNADisease[45]. After removing excess LDAs whose diseases don't arise in the 5430 known MDAs mentioned above, we would acquire 250 known LDAs. Likewise, we built an adjacency matrix $W_{dl}(i,j)$ to indicate the 250 known LDAs. Variable nl refer to the number of lncRNAs in the 250 known LDAs.

$$W_{dl}(i, j) = \begin{cases} 1, & \text{if } \ln cRNA \ l(j) \text{ is related to disease } d(i) \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

MiRNA-lncRNA interactions. The known MLIs was from starBase v2.0[46]. In the same way, we need to delete excess MLIs whose miRNAs and lncRNAs do not exist in the 5430 known MDAs and 250 known LDAs. At last, 9088 known MLIs were gotten and an adjacency matrix W_{ml} was used to refer to the 9088 MLIs.

$$W_{ml}(i, j) = \begin{cases} 1, & \text{if } \ln cRNA \ l(j) \text{ is related to miRNA } m(i) \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

MiRNA functional similarity. The scores of MFS were obtained from http://www.cuilab.cn/files/images/cuilab/misim.zip[47]. We used $FS(i,j)$ to indicate the score of MFS between miRNA $m(i)$ and miRNA $m(j)$.

Disease semantic similarity model 1 (DSS1). We put forward DSS1[48] on the basis of Directed Acyclic Graph (DAG)[49], which can be picked up according to MeSH descriptor of Category C. In the DAG = $(D, T(D), E(D))$ for disease D, all nodes are linked together from father to son using a straight line. The nodes of D and its elder can be collected into $T(D)$ and $E(D)$ referring to all the straight lines from father to son. Therefore, contribution of disease d in DAG(D) to the semantic value of disease D can be put forward.

$$\begin{cases} D_D1(d) = 1 & \text{if } d = D \\ D_D1(d) = \max\{\Delta * D_D1(d') | d' \in children \ of \ d\} & \text{if } d \neq D \end{cases} \tag{4}$$

where Δ is the semantic contribution decay factor. It is worthy of being mentioned that the value of contribution for disease D to its own semantic value is 1. The semantic value of disease D could be put forward.

$$DV1(D) = \sum_{d \in T(D)} D_D1(d) \tag{5}$$

At last, DSS1 between $d(i)$ and $d(j)$ can be described.

$$SS1(d(i), d(j)) = \frac{\sum_{t \in T(d(i)) \cap T(d(j))}(D_{d(i)}1(t) + D_{d(j)}1(t))}{DV1(d(i)) + DV1(d(j))} \tag{6}$$

Disease semantic similarity model 2 (DSS2). In the DSS2[48], due to the fact that a more specific disease d appearing in less DAGs would contribute more to the semantic value of disease D. Accordingly, the contribution made by d for the semantic value of D can be described by

$$D_D2(d) = -\log\left[\frac{the \ number \ of \ DAGs \ including \ d}{the \ number \ of \ disease}\right] \tag{7}$$

DSS2 between disease $d(i)$ and $d(j)$ can be defined as follows:

$$DV2(D) = \sum_{d \in T(D)} D_D2(d) \tag{8}$$

$$SS2(d(i), d(j)) = \frac{\sum_{t \in T(d(i)) \cap T(d(j))}(D_{d(i)}2(t) + D_{d(j)}2(t))}{DV2(d(i)) + DV2(d(j))} \tag{9}$$

Gaussian interaction profile kernel similarity. For disease $d(u)$, we used $IP(d(u))$ to refer to row vectors of line u in W_{dm} on the basis of known MDA. Through watching whether $d(u)$ is related to each miRNA, we computed GIPKS for diseases $d(u)$ and $d(v)$[50].

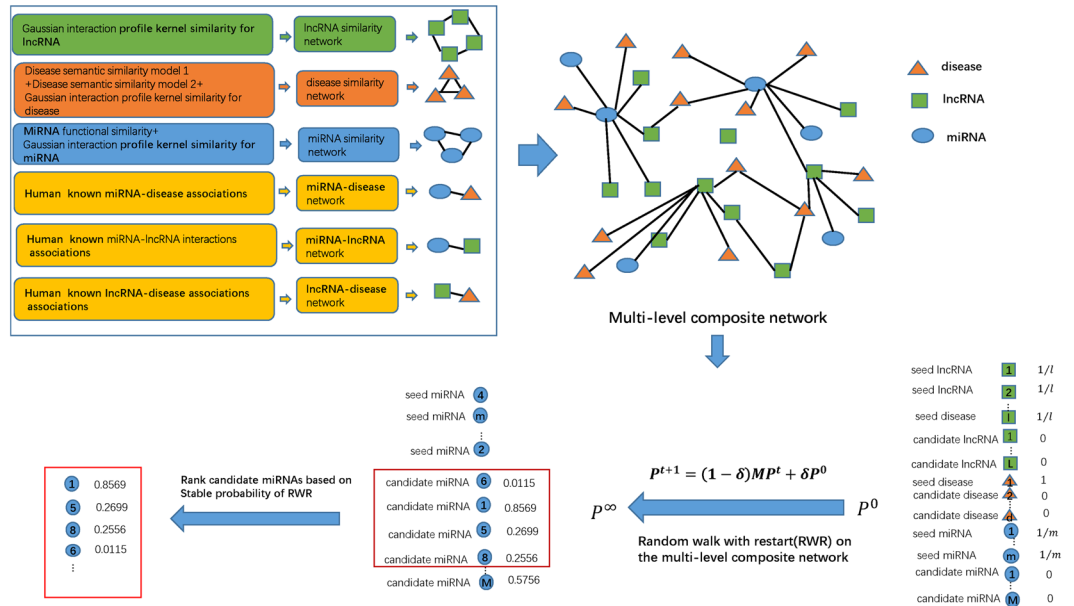$$KD(d(u), \ d(v)) = \exp(-\gamma_d \| IP(d(u)) - IP(d(v)) \|^2) \tag{10}$$

Figure 2. Flowchart of CNMDA for potential MDAs prediction in the light of HMDD v2.0. Each node in the constructed multi-level composite network possesses original probability p^0. Final scores p^∞ for MDAs would be gotten after employing RWR.

where

$$\gamma_d = \gamma'_d \Bigg/ \left(\frac{1}{nd} \sum_{u=1}^{nd} \|IP(d(u))\|^2 \right) \tag{11}$$

Similarly, GIPKS for miRNA $m(i)$ and $m(j)$ can be constructed.

$$KM(m(i),\ m(j)) = \exp(-\gamma_m \|IP(m(i)) - IP(m(j))\|^2) \tag{12}$$

where

$$\gamma_m = \gamma'_m \Bigg/ \left(\frac{1}{nm} \sum_{i=1}^{nm} \|IP(m(i))\|^2 \right) \tag{13}$$

For lncRNA $l(p)$ and $l(q)$, GIPKS between them can be constructed.

$$KL(l(p),\ l(q)) = \exp(-\gamma_l \|IP(l(p)) - IP(l(q))\|^2) \tag{14}$$

$$\gamma_l = \gamma'_l \Bigg/ \left(\frac{1}{nl} \sum_{p=1}^{nl} \|IP(l(p))\|^2 \right) \tag{15}$$

Integrated similarity for diseases (ISD) and miRNAs.   We have taken into account combining GIPKS for diseases, DSS1 and DSS2 to compute ISD between diseases $d(u)$ and $d(v)$[24].

$$SD(d(u),\ d(v)) = \begin{cases} \dfrac{SS1(d(u),\ d(v)) + SS2(d(u),\ d(v))}{2} & d(u)\ and\ d(v)\ has\ semantic\ similarity \\ KD(d(u),\ d(v)) & otherwise \end{cases} \tag{16}$$

Similarly, the ISM between miRNAs $m(i)$ and $m(j)$ can be put forward by the integration of GIPK for miRNA and MFS[24].

$$S_m(m(i),\ m(j)) = \begin{cases} FS(m(i),\ m(j)) & m(i)\ and\ m(j)\ has\ functional\ similarity \\ KM(m(i),\ m(j)) & otherwise \end{cases} \tag{17}$$

CNMDA.   Aiming at the prediction of potential MDAs, a computing method of CNMDA was stated. Carrying out RWR on a multi-level composite network that built by integration of ISM, ISD, GIPKS for lncRNA, known MDAs, LDAs and MLIs, final association scores of novel MDAs would be obtained (See Fig. 2, motivated by the

studies of Yao et al.[51]). In our introduced model, we used W_l, W_d, W_m, W_{ld}, W_{dm}, W_{lm} to indicate the initial matrix of GIPKS for lncRNAs, ISD, ISM, known LDAs, known MDAs and known MLIs, respectively. Then, the initial matrix of the multi-level composite network can be defined as $W = \begin{bmatrix} W_l & W_{ld} & W_{lm} \\ W_{ld}^T & W_d & W_{dm} \\ W_{lm}^T & W_{dm}^T & W_m \end{bmatrix}$, here, T refer to the transposition of matrix.

Global information based on the multi-level network would be captured through RWR algorithm. At each steps, seed nodes move to their immediate neighbors with a probability $(1 - \delta)$ or go back to the seed nodes with a restart probability δ. P^0 was put forward to denote the original probability vector, and P^{t+1} was introduced to represent a probability vector of node at step $t + 1$, which could be described by:

$$P^{t+1} = (1 - \delta)MP^t + \delta P^0 \tag{18}$$

where $\delta \in (0, 1)$ is a restart probability. In the multi-level network, the initial seed node probability $P^0 = \begin{bmatrix} \alpha * u_0 \\ \beta * v_0 \\ (1 - \alpha - \beta) * w_0 \end{bmatrix}$, where α, β and $(1 - \alpha - \beta)$ denote the weight of ISD network, ISM network and the network of GIPKS for lncRNAs, respectively. The corresponding u_0, v_0, w_0 are the original probabilities of these three-similarity networks respectively. Here, u_0 is calculated through assigning equal probability to all nodes in LDAs with a total to 1. Similarly, v_0, w_0 can be calculated.

Meanwhile, the transition matrix $M = \begin{bmatrix} M_l & M_{ld} & M_{lm} \\ M_{dl}^T & M_d & M_{dm} \\ M_{ml}^T & M_{md}^T & M_m \end{bmatrix}$ can be computed in the light of adjacency matrix W. $M(i,j)$ represents the transition probability from i to j. In the network of GIPKS for lncRNAs, the transition probability from lncRNA $i(l_i)$ to lncRNA $j(l_j)$ was put forward.

$$M_l(i, j) = \Pr(l_j|l_i)$$
$$= \begin{cases} (1 - x - y)W_l(i, j)/\sum_j W_l(i, j), & if \quad \sum_j W_{ld}(i, j) \neq 0 \text{ and } \sum_j W_{lm}(i, j) \neq 0 \\ (1 - x)W_l(i, j)/\sum_j W_l(i, j), & if \quad \sum_j W_{ld}(i, j) \neq 0 \text{ and } \sum_j W_{lm}(i, j) = 0 \\ (1 - y)W_l(i, j)/\sum_j W_l(i, j), & if \quad \sum_j W_{ld}(i, j) = 0 \text{ and } \sum_j W_{lm}(i, j) \neq 0 \\ W_l(i, j)/\sum_j W_l(i, j), & if \quad \sum_j W_{ld}(i, j) = 0 \text{ and } \sum_j W_{lm}(i, j) \neq 0 \end{cases} \tag{19}$$

Similarly, in the ISD network, the transition probability from disease $i(d_i)$ to disease $j(d_j)$ was put forward.

$$M_d(i, j) = \Pr(d_j|d_i)$$
$$= \begin{cases} (1 - x - z)W_d(i, j)/\sum_j W_d(i, j), & if \quad \sum_j W_{dm}(i, j) \neq 0 \text{ and } \sum_j W_{ld}(j, i) \neq 0 \\ (1 - z)W_d(i, j)/\sum_j W_d(i, j), & if \quad \sum_j W_{dm}(i, j) \neq 0 \text{ and } \sum_j W_{ld}(j, i) = 0 \\ (1 - x)W_d(i, j)/\sum_j W_d(i, j), & if \quad \sum_j W_{dm}(i, j) = 0 \text{ and } \sum_j W_{ld}(j, i) \neq 0 \\ W_d(i, j)/\sum_j W_d(i, j), & if \quad \sum_j W_{dm}(i, j) = 0 \text{ and } \sum_j W_{ld}(j, i) = 0 \end{cases} \tag{20}$$

In the ISM network, the transition probability from miRNA $i(m_i)$ to miRNA $j(m_j)$ was put forward.

$$M_m(i, j) = \Pr(m_j|m_i)$$
$$= \begin{cases} (1 - y - z)W_m(i, j)/\sum_j W_m(i, j), & if \quad \sum_j W_{dm}(j, i) \neq 0 \text{ and } \sum_j W_{lm}(j, i) \neq 0 \\ (1 - y)W_m(i, j)/\sum_j W_m(i, j), & if \quad \sum_j W_{dm}(j, i) \neq 0 \text{ and } \sum_j W_{lm}(j, i) = 0 \\ (1 - z)W_m(i, j)/\sum_j W_m(i, j), & if \quad \sum_j W_{dm}(j, i) = 0 \text{ and } \sum_j W_{lm}(j, i) \neq 0 \\ W_m(i, j)/\sum_j W_m(i, j), & if \quad \sum_j W_{dm}(j, i) = 0 \text{ and } \sum_j W_{lm}(j, i) = 0 \end{cases} \tag{21}$$

In the LDAs network, the transition probability from lncRNA $i(l_i)$ to disease $j(d_j)$ was put forward.

$$M_{ld}(i, j) = \Pr(d_j|l_i) = \begin{cases} xW_{ld}(i, j)/\sum_j W_{ld}(i, j), & if \quad \sum_j W_{ld}(i, j) \neq 0 \\ 0, & otherwise \end{cases} \tag{22}$$

In the MLIs network, transition probability from lncRNA $i(l_i)$ to miRNA $j(m_j)$ was put forward.

$$M_{lm}(i, j) = \Pr(m_j|l_i) = \begin{cases} yW_{lm}(i, j)/\sum_j W_{lm}(i, j), & if \ \sum_j W_{lm}(i, j) \neq 0 \\ 0, & otherwise \end{cases} \quad (23)$$

In the LDAs network, the transition probability from disease $i(d_i)$ to lncRNA $j(l_j)$ was put forward.

$$M_{dl}(i, j) = \Pr(l_j|d_i) = \begin{cases} xW_{ld}(j, i)/\sum_j W_{ld}(j, i), & if \ \sum_j W_{ld}(j, i) \neq 0 \\ 0, & otherwise \end{cases} \quad (24)$$

In the MDAs network, the transition probability from disease $i(d_i)$ to miRNA $j(m_j)$ was put forward.

$$M_{dm}(i, j) = \Pr(m_j|d_i) = \begin{cases} zW_{dm}(i, j)/\sum_j W_{dm}(i, j), & if \ \sum_j W_{dm}(j, i) \neq 0 \\ 0, & otherwise \end{cases} \quad (25)$$

In the MLIs network, the transition probability from miRNA $i(m_i)$ to lncRNA $j(l_j)$ was put forward.

$$M_{ml}(i, j) = \Pr(l_j|m_i) = \begin{cases} yW_{lm}(j, i)/\sum_j W_{lm}(j, i), & if \ \sum_j W_{lm}(j, i) \neq 0 \\ 0, & otherwise \end{cases} \quad (26)$$

In the MDAs network, the transition probability from miRNA $i(m_i)$ to disease $j(d_j)$ was put forward.

$$M_{md}(i, j) = \Pr(d_j|m_i) = \begin{cases} zW_{md}(j, i)/\sum_j W_{md}(j, i), & if \ \sum_j W_{md}(j, i) \neq 0 \\ 0, & otherwise \end{cases} \quad (27)$$

where x, y, z are the jumping probability between the network of GIPKS for lncRNAs and ISD network, between the network of GIPKS for lncRNAs and ISM network, and between ISD network and ISM network, respectively.

CNMDA is performed until the probabilities tend to a steady state, $P^\infty = \begin{bmatrix} \alpha * u_\infty \\ \beta * v_\infty \\ (1 - \alpha - \beta) * w_\infty \end{bmatrix}$ (the range between P^t and P^0 computed by L_1 norm is smaller than 10^{-6}). Then, the candidate miRNAs can be ranked according to w_∞.

By incorporating MLIs and LDA into MDAs prediction, RWR was put forward on a constructed multi-level network to infer novel MDAs. In the network, because initial MLIs, LDAs and MDAs have more credibility, they all as weights in the RWR equations. Obviously, the one interaction and two associations play an equally important part in the network to disseminate information of miRNAs, diseases and lncRNAs for the novel MDAs prediction. In this study, we chose the same parameter as the one in previous literature[51], which used RWR on the same multi-level composite network in their study. Therefore, we set the parameter δ to 0.7 and x, y, z, α, β to $\frac{1}{3}$.

References

1. Wienholds, E. & Plasterk, R. H. A. MicroRNA expression in zebrafish embryonic development. *Science* **309**, 310–311 (2005).
2. Cheng, A. M., Byrom, M. W., Jeffrey, S. & Ford, L. P. Antisense inhibition of human miRNAs and indications for an involvement of miRNA in cell growth and apoptosis. *Nucleic Acids Research* **33**, 1290–1297 (2005).
3. Karp, X. & Ambros, V. Encountering microRNAs in cell fate signaling. *Science* **310**, 1288–1289 (2005).
4. Alshalalfa, M. & Alhajj, R. Using context-specific effect of miRNAs to identify functional associations between miRNAs and gene signatures. *Bmc Bioinformatics* **14**, S1 (2013).
5. Miska, E. A. How microRNAs control cell division, differentiation and death. *Curr Opin Genet Dev* **15**, 563–568, https://doi.org/10.1016/j.gde.2005.08.005 (2005).
6. Bartel, D. P. MicroRNA Target Recognition and Regulatory Functions. *Cell* **136**, 215–233 (2009).
7. Xu, P., Guo, M. & Hay, B. A. MicroRNAs and the regulation of cell death. *Trends in genetics: TIG* **20**, 617–624, https://doi.org/10.1016/j.tig.2004.09.010 (2004).
8. Esquela-Kerscher, A. & Slack, F. J. Oncomirs - microRNAs with a role in cancer. *Nature reviews. Cancer* **6**, 259–269, https://doi.org/10.1038/nrc1840 (2006).
9. Meola, N., Gennarino, V. A. & Banfi, S. microRNAs and genetic diseases. *PathoGenetics* **2**, 7, https://doi.org/10.1186/1755-8417-2-7 (2009).
10. Tavazoie, S. F. *et al.* Endogenous human microRNAs that suppress breast cancer metastasis. *Nature* **451**, 147–152 (2008).
11. Slaby, O. *et al.* Altered expression of miR-21, miR-31, miR-143 and miR-145 is related to clinicopathologic features of colorectal cancer. *Oncology* **72**, 397–402 (2007).
12. Calin, G. A. & Croce, C. M. MicroRNA signatures in human cancers. *Nature Reviews Cancer* **6**, 857–866 (2006).
13. Zeng, X., Liu, L., Lu, L. & Zou, Q. Prediction of potential disease-associated microRNAs using structural perturbation method. *Bioinformatics* **34**, 2425–2432, https://doi.org/10.1093/bioinformatics/bty112 (2018).
14. Zeng, X., Zhang, X. & Zou, Q. Integrative approaches for predicting microRNA function and prioritizing disease-related microRNA using biological interaction networks. *Briefings in bioinformatics* **17**, 193–203, https://doi.org/10.1093/bib/bbv033 (2016).
15. Chen, X., Wang, L., Qu, J., Guan, N. N. & Li, J. Q. Predicting miRNA-disease association based on inductive matrix completion. *Bioinformatics (Oxford, England)*, https://doi.org/10.1093/bioinformatics/bty503 (2018).
16. Chen, X., Zhou, Z. & Zhao, Y. ELLPMDA: Ensemble learning and link prediction for miRNA-disease association prediction. *RNA Biol*, 1–12, https://doi.org/10.1080/15476286.2018.1460016 (2018).
17. Chen, X., Wu, Q. F. & Yan, G. Y. RKNNMDA: Ranking-based KNN for MiRNA-Disease Association prediction. *RNA Biol* **14**, 952–962, https://doi.org/10.1080/15476286.2017.1312226 (2017).
18. Li, J. Q., Rong, Z. H., Chen, X., Yan, G. Y. & You, Z. H. MCMDA: Matrix completion for MiRNA-disease association prediction. *Oncotarget* **8**, 21187–21199 (2017).
19. Chen, X. *et al.* RBMMMDA: predicting multiple types of disease-microRNA associations. *Scientific Reports* **5**, 13877 (2015).

20. Chen, X. *et al*. HGIMDA: Heterogeneous graph inference for miRNA-disease association prediction. *Oncotarget* **7**, 65257–65269 (2016).
21. Xuan, P. *et al*. Correction: Prediction of microRNAs Associated with Human Diseases Based on Weighted k Most Similar Neighbors. *Plos One* **8**, e70204 (2013).
22. Xuan, P. *et al*. Prediction of potential disease-associated microRNAs based on random walk. *Bioinformatics* **31**, 1805–1815 (2015).
23. Chen, X. & Yan, G. Y. Semi-supervised learning for potential human microRNA-disease associations inference. *Scientific Reports* **4**, 5501 (2014).
24. Chen, X. *et al*. WBSMDA: Within and Between Score for MiRNA-Disease Association prediction. *Scientific Reports* **6**, 21106, https://doi.org/10.1038/srep21106 (2016).
25. Jiang, Q. *et al*. Prioritization of disease microRNAs through a human phenome-microRNAome network. *Bmc Systems Biology* **4**(Suppl 1), S2 (2010).
26. Mork, S., Pletscher-Frankild, S., Palleja Caro, A., Gorodkin, J. & Jensen, L. J. Protein-driven inference of miRNA-disease associations. *Bioinformatics* **30**, 392–397, https://doi.org/10.1093/bioinformatics/btt677 (2014).
27. Shi, H. *et al*. Walking the interactome to identify human miRNA-disease associations through the functional link between miRNA targets and disease genes. *Bmc Systems Biology* **7**, 1–12 (2013).
28. Chen, X., Liu, M. X. & Yan, G. Y. RWRMDA: predicting novel human microRNA-disease associations. *Mol Biosyst* **8**, 2792–2798, https://doi.org/10.1039/c2mb25180a (2012).
29. Li, Y. *et al*. HMDDv2.0: a database for experimentally supported human microRNA and disease associations. *Nucleic Acids Res* **42**, D1070–1074, https://doi.org/10.1093/nar/gkt1023 (2014).
30. Lu, M. *et al*. An analysis of human microRNA and disease associations. *Plos One* **3**, e3420 (2008).
31. Linehan, W. M., Grubb, R. L., Coleman, J. A., Zbar, B. & Walther, M. M. The genetic basis of cancer of kidney cancer: implications for gene-specific clinical management. *Bju International* **95**, 2–7 (2005).
32. Sudarshan, S. & Linehan, W. M. Genetic basis of cancer of the kidney. *Seminars in oncology* **33**, 544–551, https://doi.org/10.1053/j.seminoncol.2006.06.008 (2006).
33. Lamm, D. L. Cancer statistics. *CA: a cancer journal for clinicians* **40**, 318–319 (1990).
34. Zaman, M. S. *et al*. Up-Regulation of MicroRNA-21 Correlates with Lower Kidney Cancer Survival. *Plos One* **7**, e31060–e31060 (2012).
35. Wu, D. *et al*. microRNA-133b downregulation and inhibition of cell proliferation, migration and invasion by targeting matrix metallopeptidase-9 in renal cell carcinoma. *Molecular Medicine Reports* **10**, 2491–2498 (2014).
36. Smigal, C. *et al*. Trends in breast cancer by race and ethnicity: update 2006. *CA: a cancer journal for clinicians* **56**, 168–183 (2006).
37. Group, E. B. C. T. C. Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *Lancet* **365**, 1687–1717 (2005).
38. Heneghan, H. M., Miller, N., Kelly, R., Newell, J. & Kerin, M. J. Systemic miRNA-195 Differentiates Breast Cancer from Other Malignancies and Is a Potential Biomarker for Detecting Noninvasive and Early Stage Disease. *Oncologist* **15**, 673–682 (2010).
39. Yang, Y. *et al*. The role of microRNA in human lung squamous cell carcinoma. *Cancer Genetics & Cytogenetics* **200**, 127–133 (2010).
40. Yanaihara, N. *et al*. Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell* **9**, 189–198 (2006).
41. Chen, X., Xie, D., Zhao, Q. & You, Z. H. MicroRNAs and complex diseases: from experimental results to computational models. *Briefings in bioinformatics*, https://doi.org/10.1093/bib/bbx130 (2017).
42. You, Z. H. *et al*. PBMDA: A novel and effective path-based computational model for miRNA-disease association prediction. *Plos Computational Biology* **13**, e1005455 (2017).
43. Chen, X., Yan, C. C., Zhang, X. & You, Z. H. Long non-coding RNAs and complex diseases: from experimental results to computational models. *Briefings in bioinformatics* **18**, 558–576, https://doi.org/10.1093/bib/bbw060 (2017).
44. Chen, X. *et al*. Drug-target interaction prediction: databases, web servers and computational models. *Briefings in bioinformatics* **17**, 696–712, https://doi.org/10.1093/bib/bbv066 (2016).
45. Chen, G. *et al*. LncRNADisease: a database for long-non-coding RNA-associated diseases. *Nucleic Acids Res* **41**, D983–986, https://doi.org/10.1093/nar/gks1099 (2013).
46. Li, J. H., Liu, S., Zhou, H., Qu, L. H. & Yang, J. H. StarBasev2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res* **42**, D92–97, https://doi.org/10.1093/nar/gkt1248 (2014).
47. 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* **26**, 1644–1650, https://doi.org/10.1093/bioinformatics/btq241 (2010).
48. Xuan, P. *et al*. Prediction of microRNAs associated with human diseases based on weighted k most similar neighbors. *PLoS One* **8**, e70204, https://doi.org/10.1371/journal.pone.0070204 (2013).
49. Chen, X., Huang, Y. A., Wang, X. S., You, Z. H. & Chan, K. C. FMLNCSIM: fuzzy measure-based lncRNA functional similarity calculation model. *Oncotarget* **7**, 45948–45958, https://doi.org/10.18632/oncotarget.10008 (2016).
50. van Laarhoven, T., Nabuurs, S. B. & Marchiori, E. Gaussian interaction profile kernels for predicting drug-target interaction. *Bioinformatics* **27**, 3036–3043, https://doi.org/10.1093/bioinformatics/btr500 (2011).
51. Yao, Q. *et al*. Global Prioritizing Disease Candidate lncRNAs via a Multi-level Composite Network. *Sci Rep* **7**, 39516, https://doi.org/10.1038/srep39516 (2017).

Author Contributions

B.S.H. put forward the subject and its prediction approach, also, wrote the manuscript. J.Q. carried out the experiments and amended this manuscript. M.C. analyzed the outcome and amended manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-34180-6.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.