SCIENTIFIC REPORTS

OPEN

Inferring microRNA-disease association by hybrid recommendation algorithm and unbalanced bi-random walk on heterogeneous network

Dong-Ling Yu[1], Yuan-Lin Ma[1] & Zu-Guo Yu [1,2]

More and more research works have indicated that microRNAs (miRNAs) play indispensable roles in exploring the pathogenesis of diseases. Detecting miRNA-disease associations by experimental techniques in biology is expensive and time-consuming. Hence, it is important to propose reliable and accurate computational methods to exploring potential miRNAs related diseases. In our work, we develop a novel method (BRWHNHA) to uncover potential miRNAs associated with diseases based on hybrid recommendation algorithm and unbalanced bi-random walk. We first integrate the Gaussian interaction profile kernel similarity into the miRNA functional similarity network and the disease semantic similarity network. Then we calculate the transition probability matrix of bipartite network by using hybrid recommendation algorithm. Finally, we adopt unbalanced bi-random walk on the heterogeneous network to infer undiscovered miRNA-disease relationships. We tested BRWHNHA on 22 diseases based on five-fold cross-validation and achieves reliable performance with average AUC of 0.857, which an area under the ROC curve ranging from 0.807 to 0.924. As a result, BRWHNHA significantly improves the performance of inferring potential miRNA-disease association compared with previous methods. Moreover, the case studies on lung neoplasms and prostate neoplasms also illustrate that BRWHNHA is superior to previous prediction methods and is more advantageous in exploring potential miRNAs related diseases. All source codes can be downloaded from https://github.com/myl446/BRWHNHA.

MicroRNAs (miRNAs) are a class of short non–coding RNAs (21–25 nt)[1–3]. As an important transcriptional regulatory factor, miRNAs are widely involved in the biological procedures of disease-related gene regulation, which is closely related to human multi-gene diseases[4–7]. Increasing evidences have demonstrated that miRNAs play a critical role in the emergence and development of diseases[8,9]. Hence, revealing miRNAs associated diseases is an efficient way to accelerate the acquaintance about disease pathology at the molecular level[10–13].

As detecting miRNA-disease associations by experimental techniques is expensive and time-consuming, many effective computational methods about the prediction of the relationship between miRNAs and diseases have been proposed. For example, Jiang et al.[14] proposed a computational approach to infer potential miRNA-disease associations by hypergeometric distribution. For a given disease, they priorited the entire human miRNAs. In addition, Jiang et al.[15] further improved the calculation of concordance score between a miRNA and a given disease. Chen et al.[16] firstly presented a prediction computational method named RWRMDA based on global network similarity, to predict novel human miRNA-disease associations by adopting the method of random walk on network of miRNA functional similarity. Then, Xuan et al.[17] developed a reliable prediction method based on random walk, they assigned different weights to transition matrix of miRNAs depending on whether they are

[1]Key Laboratory of Intelligent Computing and Information Processing of Ministry of Education and Hunan Key Laboratory for Computation and Simulation in Science and Engineering, Xiangtan University, Xiangtan, Hunan 411105, P.R. China. [2]School of Electrical Engineering and Computer Science, Queensland University of Technology, Brisbane, Q4001, Australia. Yuan-Lin Ma and Dong-Ling Yu contributed equally. Correspondence and requests for materials should be addressed to Z.-G.Y. (email: yuzuguo@aliyun.com)

associated with given diseases to exploit the prior information of nodes and the various ranges of topologies. And they extended the walk on a miRNA-disease bipartite network to predict candidates miRNAs, specially for the diseases without any known related miRNAs. Furthermore, Chen et al.[18] developed a novel prediction method named WBSMDA for inferring miRNA-disease based on integrating miRNA functional similarity, disease semantic similarity, the known miRNA-disease associations, and the Gaussian interaction profile kernel similarity into heterogeneous network. WBSMDA not only could deal with new diseases without any known associated miRNAs, but also could handle new miRNAs without any known associated disease. In 2016, Zeng et al.[19] conducted a review on methods for predicting disease and miRNA associations based on biological interaction networks. After detailed comparing these methods, they pointed out the current challenges in predicting disease and miRNA correlations. Liu et al.[20] further proposed a method to explore potential miRNAs related to diseases by integrating multiple biology data in 2017. In recent years, recommendation system algorithms have been successfully applied in many fields. Chen et al.[21] presented a hybrid approach for miRNA-disease association prediction (HAMDA) method based on hybrid recommendation methods, which combined available biology data and network-based inference methods. However, just like the above mentioned methods, they only prioritized miRNAs by utilizing the same layers neighbor nodes of miRNAs and diseases rather than making use of the different structural and topological characteristics among subgraphs of heterogeneous networks. Luo et al.[22] proposed a novel effective prediction model that use unbalanced bi-random walk to improve performance of prediction. They fully exploited the different topological and structural of miRNA similarity networks and disease similarity network. This method improved prediction performance, but ignored the prior information and the respective topological structural of bipartite network. Zeng et al.[23] found that heterogeneous miRNA-disease networks perform better on prediction than single disease similarity networks, miRNA similarity networks, and the known disease-gene association networks in 2018. So they adopted a method of structural perturbation to improve the prediction accuracy of miRNA-disease association.

We believe that the topological and structural features of heterogeneous network contain important information which is useful for discovering more reliable miRNA-disease associations. In present work, we develop an efficient computational method based on hybrid recommendation approach and unbalanced bi-random walk, called BRWHNHA (Bi-random Walk on Heterogeneous Network based on Hybrid Approach), which exploits the characteristic of nodes and the topological structural of the known miRNA-disease association by using hybrid recommendation approach, and taken advantage of the different topological structural between similarity networks of miRNA and disease by adopting bi-random walk on heterogeneous network. The hybrid recommendation algorithm adds some virtual edges to heterogeneous networks by calculating the transition matrix of bipartite network, so that the unbalanced bi-random walk on the new heterogeneous network can find potential miRNAs related to diseases more efficiently. To validate the prediction ability of BRWHNHA, we adopted five-fold cross-validation and compared BRWHNHA with MIDPE[17], HAMDA[21], and BRWH[22]. The average AUC is 2.13%, 0.69%, and 2.20% higher than the three methods. The case studies on lung neoplasms and prostatic neoplasms, and in the top 50 predicted associations, there are 49 and 46 real associations, respectively. It further demonstrates the ability of BRWHNHA in discovering potential miRNAs associated with disease.

Results

To evaluate the prediction effectiveness of BRWHNHA in exploring undiscovered association between miRNAs and diseases, we compared BRWHNHA with MIDPE[17], HAMDA[21], and BRWH[22] by five-fold cross-validation with repeating 100 times on the dataset obtained by Luo and Xiao[22]. For a given disease, we randomly divided the known-related miRNAs into five subsets with equal size. For each round, we used one subset as testing set and other four subsets as training set. After 5 rounds, we calculated the average AUC value. In order to reduce false positive, we recalculated miRNAs similarity and obtained a bran-new similarity matrix in each round of prediction. Then we calculated the probability of association between the given disease and miRNAs by BRWHNHA. Finally, all candidate miRNAs were ranked by association probability. The higher the miRNAs in testing set were ranked, the better the performance. As the most of diseases only have a few association with miRNAs that have been proved, the performance of the prediction methods can not be accurately evaluated. Hence we only tested the 22 diseases associated with at least 60 miRNAs as Luo and Xiao[22]. We only showed recall-precision curve of breast neoplasms and lung neoplasms. In addition, we analyzed effect of parameters on performance of BRWHNHA.

Performance evaluation.   In this study, the novelty of BRWHNHA was to calculate the transition probability matrix of bipartite network by using hybrid recommendation algorithm, and then a bi-random walk on heterogeneous network based on hybrid approach was adopted. The average AUC is 83.55% without using hybrid recommendation algorithm, which was 2.14% less than using hybrid recommendation algorithm. Therefore, it is important to construct the transition probability matrix by hybrid recommendation algorithm. The prediction accuracy was actually improved by exploring the prior information and topological structure of bipartite networks. The same heterogeneous network was used on MIDPE[17], HAMDA[21] and BRWH[22]. The best parameters of $\alpha = 0.9$ and $\gamma = 0.8$ for MIDPE, $\sigma = 0.7$ and $\rho = 0.8$ for HAMDA $\lambda = 0.6$, $\alpha = 0.4$, $r = 2$, $l = 1$ for BRWH were adopted as reported in original papers.

As illustrated in Table 1, the average AUC values of MIDPE, HAMDA, BRWHA and BRWHNHA in 22 diseases are 83.55%, 85.00%, 83.49% and 85.69% respectively. BRWHNHA performed the best with AUC 2.13%, 0.69% and 2.20% higher than other three methods. Moreover, BRWHNHA is superior to MIDPE and BRWH in all measurements for 22 diseases. Although HAMDA achieves higher AUC than BRWHNHA in 7 out of 22 diseases, but BRWHNHA obtains better performance in most of diseases. Since HAMDA repeatedly uses the known miRNAs-disease association data in the measurement of miRNAs similarity, it maybe overestimate the results.

Diseases name	Number of related miRNAs	The average AUC			
		MIDPE[17]	HAMDA[21]	BRWH[22]	BRWHNHA
Breast Neoplasms	202	0.813	0.821	0.812	**0.855**
Carcinoma, Hepatocellular	214	0.777	0.791	0.776	**0.805**
Carcinoma, Non-Small-Cell Lung	95	0.859	0.867	0.857	**0.874**
Carcinoma, Renal Cell	107	0.814	0.822	0.812	**0.828**
Carcinoma, Squamous Cell	80	0.874	**0.886**	0.875	0.882
Colonic Neoplasms	78	0.855	0.866	0.853	**0.872**
Colorectal Neoplasms	147	0.819	0.838	0.824	**0.848**
Endometriosis	62	0.815	**0.850**	0.814	0.841
Esophageal Neoplasms	74	0.794	0.793	0.793	**0.825**
Glioblastoma	96	0.805	0.825	0.801	**0.839**
Glioma	71	0.868	**0.878**	0.863	0.874
Head and Neck Neoplasms	64	0.872	**0.886**	0.870	0.881
Heart Failure	120	0.802	**0.810**	0.800	0.807
Leukemia, Myeloid, Acute	64	0.857	**0.868**	0.852	0.858
Lung Neoplasms	132	0.906	0.919	0.906	**0.924**
Medulloblastoma	62	0.801	0.807	0.798	**0.811**
Melanoma	141	0.823	0.838	0.825	**0.849**
Ovarian Neoplasms	114	0.866	0.902	0.887	**0.903**
Pancreatic Neoplasms	99	0.905	0.907	0.902	**0.911**
Prostatic Neoplasms	118	0.833	0.857	0.831	**0.864**
Stomach Neoplasms	173	0.782	0.802	0.783	**0.832**
Urinary Bladder Neoplasms	92	0.842	**0.867**	0.833	**0.867**

Table 1. Predicting outcomes for MIDPE, HAMDA, BRWH and BRWHNHA by the five-fold cross-validation.
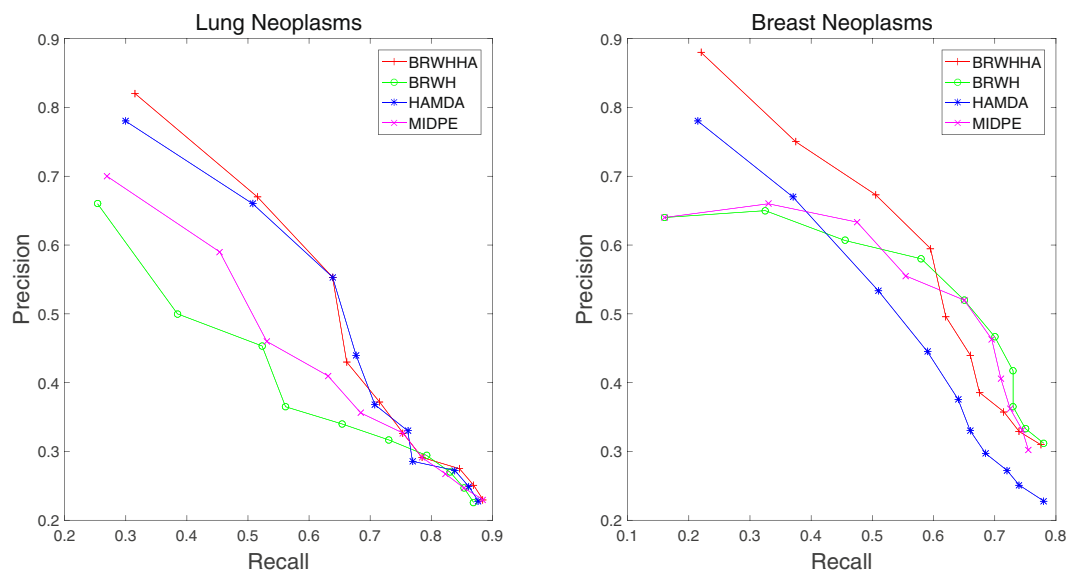

Figure 1. Recall-precision curves of breast neoplasm and lung neoplasm by five-fold cross-validation.

ROC curves of BRWHNHA and other three methods corresponding to the maximum AUC value in five-fold cross-validation at 100 times have shown in Supplementary Fig. S1 in Additional file.

In Fig. 1, we compared BRWHNHA with other three methods in the recall-precision curves of breast neoplasms and lung neoplasms based on five-fold cross-validation. The precision-recall curve was obtained by measuring recall and precision at positions of top k ($k = 10, 20, …, 100$). The results show that our method achieves the highest precision and recall in the top 20. Moreover, with the increase of k value, the precision of BRWHNHA decreases, but the recall increases. It suggests that the associations ranked in top position have higher probability of being potential miRNA-disease associations. We also compared the statistical significance of the difference in predictive ability between BRWHNHA and other three methods by paired t-tests. The P-values are listed in

Method	MIDPE[17]	HAMDA[21]	BRWH[22]
P-value	2.41E-07	1.71E-02	4.00E-08

Table 2. Pairwise comparison between BRWHNHA and another method by paired t-test on the AUC of prediction.

$\alpha = 0.4, r = 2, l = 1$		$\lambda = 0.6, r = 2, l = 1$	
λ	Average AUC	α	Average AUC
0	0.748654	0	0.804929
0.1	0.814260	0.1	0.844848
0.2	0.844524	0.2	0.853413
0.3	0.852676	0.3	0.856242
0.4	0.855220	0.4	**0.856818**
0.5	0.856081	0.5	0.855095
0.6	**0.856818**	0.6	0.853919
0.7	0.856418	0.7	0.850477
0.8	0.856146	0.8	0.848039
0.9	0.855824	0.9	0.847697
1	0.855415	1	0.847478
$\lambda = 0.6, \alpha = 0.4, l = 1$		$\lambda = 0.6, \alpha = 0.4, r = 2$	
r	Average AUC	l	Average AUC
0	0.846267	0	0.856710
1	0.852731	1	**0.856818**
2	**0.856818**	2	0.853561
3	0.856814	3	0.846865
4	0.856555	4	0.846626
5	0.856371	5	0.846505

Table 3. Effects of parameters λ, α, r, l on prediction performance of BRWHNHA.

Table 2. Obviously, BRWHNHA achieves better performance than MIDPE, HAMDA, BRWHA at the significance level of 0.05.

We also compared our method with SPM on the dataset used by Zeng et al.[23]. We found that our method performs slightly better than SPM on five subsets with equal size in five-fold cross-validation in most cases (comparison results are not shown here).

Effect of parameters in BRWHNHA. There are four parameters λ, α, r and l explored in our method. The parameter λ is the hybridization parameter to mediate between HeatS algorithm and ProbS algorithm different kinds of resource distribution processes, and parameter α plays the role to control the consistence between the predicted candidate miRNA-disease associations and the known associations. The parameters of r and l are the numbers of maximal random walk steps in miRNA similarity network and disease similarity network, respectively. We set various values of λ and α ranging from 0 to 1, the step length was 0.1. r and l were taken to be between 0 step to 5 steps, the step length was 1. Then, we calculated average AUC in the framework of five-fold cross-validation. Table 3 shows the effects of λ, α, r and l on the cross validation result in miRNA-disease association dataset. It can be observed that BRWHNHA achieves the best performance, when $\lambda = 0.6$, $\alpha = 0.4$, $r = 2$, $l = 1$.

Case study

To further validate efficiency of BRWHNHA for discovering the potential associations between miRNAs and diseases, we conducted two case studies of Lung neoplasms and Prostatic neoplasms here. All known miRNA-disease associations released in June 2014 were regarded as training sets, and the set of candidate associations formed by all other associations. The prediction results of Lung neoplasms and Prostatic neoplasms were confirmed based on relevant literatures and two important public database: dbDEMC[24] and MiR2Disease[25].

Lung cancer is one of the malignant tumors with the highest morbidity and mortality, and it is the greatest health and life threat to human. Over the past 50 years, many countries have reported significant increases in the incidence and mortality of lung cancer. The first 50 predicted miRNA associated with lung cancer were shown in Table 4. As a result, among the top 20 and 50 potential Lung neoplasms associated miRNAs, 20 and 49 were confirmed by dbDEMC database, MiR2Disease and literature. Though there is no database or literature that proved the miRNA (hsa-mir-200) relevance to lung neoplasms, the mir-200 family, which includes 5 members (miRNA-200a, miRNA-200b, miRNA-200c, miRNA-429, and miRNA-141), is associated with Lung neoplasms in dbDEMC, so we have reason to believe that it is related to the disease. In addition, we also listed the potential miRNAs from top 51 to top100 (Supplementary Table S2 in Additional file 1).

Rank	miRNAs	Evidence	Rank	miRNAs	Evidence
1	hsa-mir-20b	DB	26	hsa-mir-302f	DB
2	hsa-mir-15b	DB	27	hsa-mir-1258	DB
3	hsa-mir-34b	DB	28	hsa-mir-1305	DB
4	hsa-mir-21	DB, MD	29	hsa-mir-140	DB, MD
5	hsa-mir-200b	DB, MD	30	hsa-mir-106a	DB, MD
6	hsa-mir-29b	DB, MD	31	hsa-mir-219	DB, MD
7	hsa-mir-146a	DB, MD	32	hsa-mir-1827	PMID:21676885
8	hsa-mir-7i	DB	33	hsa-mir-675	DB
9	hsa-mir-1236	DB	34	hsa-mir-485	DB
10	hsa-mir-30e	DB, MD	35	hsa-mir-105	DB
11	hsa-let-200a	DB, MD	36	hsa-mir-92b	DB
12	hsa-mir-885	DB	37	hsa-mir-1323	DB
13	hsa-let-147b	DB	38	hsa-mir-135a	DB
14	hsa-let-10a	DB	39	hsa-mir-98	DB, MD
15	hsa-mir-198	DB, MD	40	hsa-mir-137	DB
16	hsa-let-100	DB	41	hsa-mir-27a	DB
17	hsa-mir-212	DB, MD	42	hsa-mir-235	DB
18	hsa-mir-181d	DB	43	hsa-mir-450b	DB
19	hsa-mir-217	DB	44	hsa-mir-495	DB
20	hsa-mir-204	DB, MD	45	hsa-mir-18a	DB, MD
21	hsa-mir-374b	DB	46	hsa-mir-1915	DB
22	hsa-mir-133a	DB	47	hsa-mir-101	DB, MD
23	hsa-mir-200	Unconfirmed	48	hsa-mir-23a	DB
24	hsa-mir-106b	DB	49	hsa-mir-1207	DB
25	hsa-mir-194	DB	50	hsa-mir-337	DB

Table 4. The first 50 potential miRNAs associated with lung neoplasms predicted by BRWHNHA. *The databases dbDEMC and mir2disease are respectively represented as DB and MD.

Prostate neoplasms is an important malignant tumor in male patients. There are usually no clinical symptoms in the early stage. Currently, most of the patients admitted by prostate neoplasms are in the late stage[26]. Therefore, early diagnosis is an urgent problem. There are many evidences that have confirmed a link between miRNA and prostate neoplasms, and it could be therapeutically useful for the treatment of prostate neoplasms by regulating the expression of related miRNAs[27,28]. As a result of the case study for prostate neoplasms, 18 out of the top-20 and 46 out of the top-50 predicted miRNAs of prostate neoplasms were verified by dbDEMC, MiR2Disease and literature (shown in Table 5). However, hsa-mir-302f, hsa-mir-1915, hsa-mir-4257 and hsa-mir-1286 are not included in dbDEMC, MiR2Disease and literature. We also listed the potential miRNAs from top 51 to top100 (Supplementary Table S3 in Additional file 1).

Conclusion

Taking full account of the different topological and structural characteristics of heterogeneous network is a very challenging and meaningful task in prioritizing potential disease-related miRNAs. In this paper, we first adopted an effective measurement, which is suitable for miRNAs and diseases without known miRNA-disease associations, to estimate the similarity of miRNAs and diseases. Then, we presented a BRWHNHA method based on hybrid recommendation algorithm and unbalanced bi-random walk to predict potential diseases associated miRNAs. We made full use of the prior information and topological structural by calculating the transition probability matrix of bipartite network in using hybrid recommendation algorithm, in addition, we fully exploited the topologies and structures of miRNA similarity network(MMS) and disease similarity network(DDS) in the different lever by adopting unbalanced bi-random walk on heterogeneous network. To assess the performance of BRWHNHA, we compared BRWHNHA with MIDP, HAMDA and BRWH on the dataset obtained by Luo and Xiao[22]. The results indicate that BRWHNHA has the best prediction ability among these methods, the average AUC was 2.13%, 0.69% and 2.20% higher than MIDP, HAMDA and BRWH, respectively. Furthermore, case studies on lung neoplasms and prostatic were employed to further identify the performance evaluation of BRWHNHA, which the top 49 out of 50 and 46 out of 50 predicted miRNA-disease associations respectively were confirmed by recently published literature and databases of dbDEMC and MiR2Disease. The results show that BRWHNHA can be used as an effective and important method to explore the potential association between miRNAs and diseases.

Nevertheless, there is a limitation on our BRWHNHA that should be improved in future study. That is there are many parameters need to be set in this method. So a more effective method need be adopted to find the optimal parameters.

Rank	miRNAs	Evidence	Rank	miRNAs	Evidence
1	hsa-mir-154	DB	26	hsa-mir-137	PMID:26461474
2	hsa-mir-189	DB	27	hsa-mir-134	DB
3	hsa-mir-19a	DB	28	hsa-mir-30a	DB, MD
4	hsa-mir-29b	DB, MD	29	hsa-mir-138	PMID:28741117
5	hsa-mir-21	DB, MD	30	hsa-mir-302a	DB
6	hsa-mir-199b	DB, MD	31	hsa-mir-450b	DB
7	hsa-mir-141	DB, MD	32	hsa-mir-495	DB
8	hsa-mir-10a	DB, MD	33	hsa-mir-26a	DB, MD
9	hsa-mir-181	DB	34	hsa-mir-20a	DB, MD
10	hsa-mir-885	DB	35	hsa-mir-139	DB
11	hsa-let-7d	DB, MD	36	hsa-mir-107	DB
12	hsa-mir-149	DB, MD	37	hsa-mir-148b	DB
13	hsa-let-7e	DB	38	hsa-mir-2355	DB
14	hsa-let-7f	DB, MD	39	hsa-mir-302b	DB
15	hsa-mir-195	DB, MD	40	hsa-mir-337	DB
16	hsa-let-7g	DB, MD	41	hsa-mir-624	DB
17	hsa-mir-302f	Unconfirmed	42	hsa-mir-214	DB, MD
18	hsa-mir-23b	DB, MD	43	hsa-mir-205	DB, MD
19	hsa-mir-199a	DB, MD	44	hsa-mir-1286	Unconfirmed
20	hsa-mir-1915	Unconfirmed	45	hsa-mir-340	DB
21	hsa-mir-1258	DB	46	hsa-mir-1275	DB
22	hsa-mir-675	DB	47	hsa-mir-136	DB
23	hsa-mir-4257	Unconfirmed	48	hsa-mir-218	DB, MD
24	hsa-mir-18a	DB	49	hsa-mir-498	DB, MD
25	hsa-mir-101	DB, MD	50	hsa-mir-301b	PMID:29744254

Table 5. The first 50 potential miRNAs associated with prostate neoplasms predicted by BRWHNHA. *The databases dbDEMC and mir2disease are respectively represented as DB and MD.
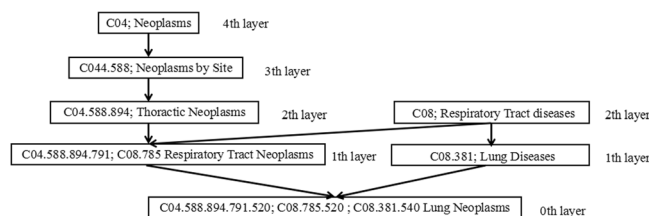
Figure 2. Hierarchical DAG graph of Lung neoplasms.

Methods

The measurement of disease semantic similarity and miRNA functional similarity. As described in the category C of MeSH descriptor, the disease relationships can be regarded as a directed acyclic graph structure. The disease K can be represented as $DAG(K) = (K, T(K), E(K))$[22], where $T(K)$ represents the set of all the ancestor nodes of disease K and disease K, $E(K)$ represents the set of all direct edges from parent nodes to child nodes in the subgraph, as shown in Fig. 2. For two diseases d_i and d_j, the disease semantic similarity measurement $DSS(d_i, d_j)$ is defined by Luo and Xiao[22].

Based on the assumption that miRNAs with similar functions are more likely to be associated with similar diseases and vice versa[29,30], the miRNA function similarity measurement $MFS(m_i, m_j)$ for two miRNAs m_i and m_j is adopted, which proposed by Wang et al.[31].

Gaussian interaction profile kernel similarity. The Gaussian interaction profile kernel similarity also is based on the assumption that miRNAs with similar functions are more likely to be associated with similar diseases and vice versa. Let $A = (a_{ij})_{n_m \times n_d}$ be the adjacency matrix of MD, and n_m denotes the number of miRNAs and n_d the number of diseases, respectively. The Gaussian interaction profile kernel similarity is calculated by the known miRNA-disease associations[21], so let $IP(d_i)$ binary vector indicate whether disease d_i is associated with each miRNA, in other words, $IP(d_i)$ is the ith column of A. The Gaussian interaction profile kernel similarity between two diseases d_i and d_j is calculated as:

$$DGS(d_i, d_j) = exp(-r_d\|IP(d_i) - IP(d_j)\|^2) \tag{1}$$

where $r_d = r'_d / \left(\frac{1}{n_d}\sum_{i=1}^{n_d}\|IP(d_i)\|^2\right)$ is the kernel bandwidth and r'_d is a new bandwidth parameter (e.g. ($r'_d = 1$ as[32,33]) to normalize r_d.

For two miRNAs m_i and m_j, the Gaussian interaction profile kernel similarity is calculated as:

$$MGS(m_i, m_j) = exp(-r_m\|IP(m_i) - IP(m_j)\|^2) \tag{2}$$

where $r_m = r'_m / \left(\frac{1}{n_m}\sum_{i=1}^{n_m}\|IP(m_i)\|^2\right)$ is the kernel bandwidth and r'_m is a new bandwidth parameter to normalize r_m.

Integrated similarity for miRNAs and diseases. A new disease similarity matrix can be obtained by integrated disease semantic similarity and the disease Gaussian interaction profile kernel similarity[21]. For two diseases d_i and d_j, the new diseases similarity can be defined as follows:

$$D_S(d_i, d_j) = \begin{cases} DSS(d_i, d_j), & DSS(d_i, d_j) \neq 0 \\ DGS(d_i, d_j), & \text{otherwise} \end{cases} \tag{3}$$

The integrated similarity between miRNAs m_i and m_j can be defined as follows:

$$M_S(m_i, m_j) = \begin{cases} MFS(m_i, m_j), & MFS(m_i, m_j) \neq 0 \\ MGS(m_i, m_j), & \text{otherwise} \end{cases} \tag{4}$$

Hybrid recommendation algorithm. A binary network $MD(M, D, E)$ is constructed by experimentally confirmed miRNA-disease association, where D represents all diseases nodes, M represents all miRNAs nodes, and E represents all edges in MD. The adjacency matrix A is defined as follows:

$$a_{ij} = \begin{cases} 1, & m_i \text{ is associated with } d_j \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

Zhou et al.[34] proposed the hybrid recommendation algorithm, which combined the heat spreading (HeatS) algorithm and probabilistic spreading (ProbS) algorithm by incorporating the hybridization parameter λ to balance the accuracy of HeatS and the diversity of ProbS. For a given disease, HeatS and ProbS both work by assigning miRNA an initial resource represented by the vector f (where f_i is the resource possessed by miRNA m_i), which was redistributed though the transformation $f' = W * f$. The miRNA that possess more resource is more likely associated the given disease. In Fig. 3, the visualization process of HeatS algorithm and ProbS algorithm is presented.

HeatS is defined as follows:

$$W_{ij}^H = \frac{1}{k(m_i)}\sum_{l=1}^{n_d}\frac{a_{il}a_{jl}}{k(d_l)} \tag{6}$$

$$f' = W^H * f \tag{7}$$

ProbS is defined as follows:

$$W_{ij}^P = \frac{1}{k(m_j)}\sum_{l=1}^{n_d}\frac{a_{il}a_{jl}}{k(d_l)} \tag{8}$$

$$f' = W^P * f \tag{9}$$

Hybrid recommendation algorithm is defined as follows:

$$W_{ij}^{H+P} = \frac{1}{k(m_i)^{1-\lambda}k(m_j)^{\lambda}}\sum_{l=1}^{nd}\frac{a_{il}a_{jl}}{k(d_l)} \tag{10}$$

$$f' = W^{H+P} * f \tag{11}$$

$k(x)$ denotes the degree of nodes x in bipartite graph $MD(M, D, E)$.

Our method BRWHNHA. In this paper, we present a BRWHNHA method based on hybrid recommendation algorithm and unbalance bi-random walk to predict potential diseases associated miRNAs. Luo et al.[22] found that most of the nodes in DDS and MMS are isolated, and the sparsity of disease semantic similarity and miRNA functional similarity effect the prediction performance. To overcome this disadvantages in data, the similarity is estimated for each disease pair via integrating disease semantic similarity and disease Gaussian interaction profile
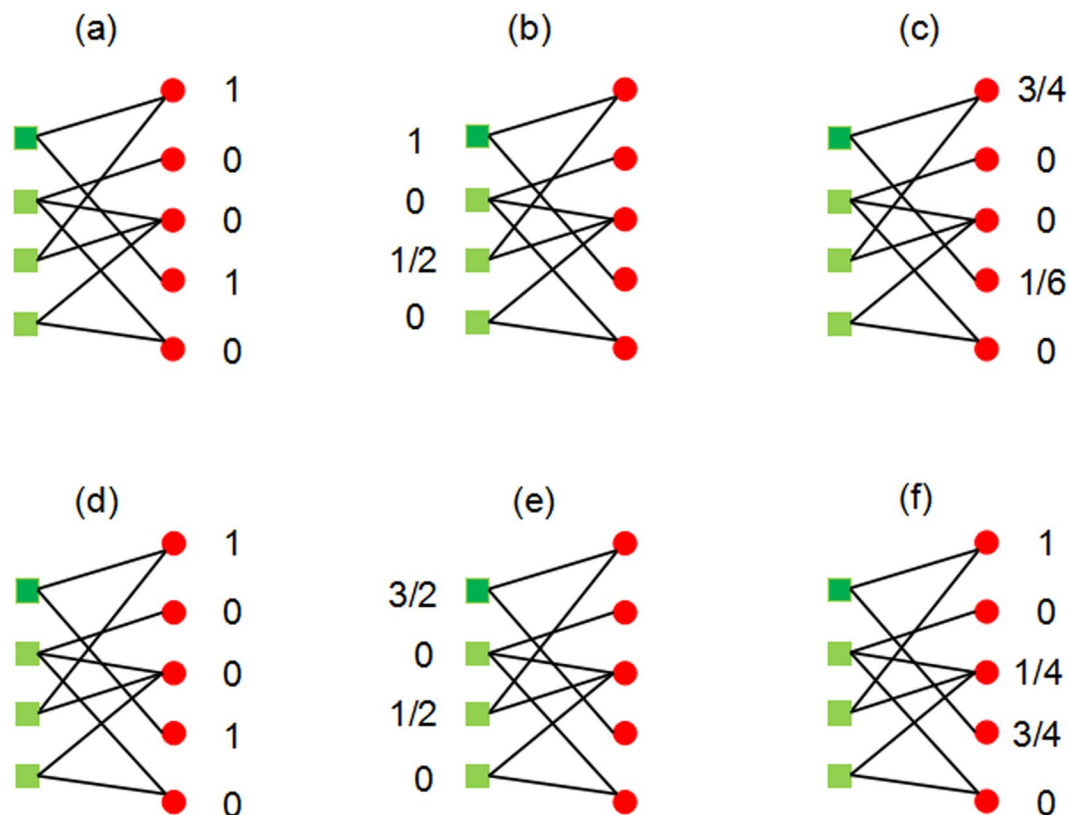
Figure 3. The HeatS (**a**–**c**) and ProbS (**d**–**f**) algorithms at work on the bipartite miRNA-disease network. Disease are shown as green squares and dark green squares is a given disease, miRNAs are shown as red circles.

kernel similarity, as well as miRNA pair is estimated via integrating miRNA function similarity and miRNA Gaussian interaction profile kernel similarity. Then, the bipartite miRNA-disease network (MD) is constructed, where edges in the miRNA-disease network are the known associations between miRNAs and diseases that were released by HMDD in June 2014. The transition probability matrix of MD is obtained by using hybrid recommendation algorithm in bipartite networks. Then, unbalance bi-random walk is carried out in heterogeneous network that includes DDS, MMS and MD. Finally, for a given disease, all candidate miRNAs will be ranked according to transition probability matrix, and the higher the rank, the more likely it is to be associated with the given disease. Flowchart of potential miRNA-disease association prediction based on the computational model of BRWHNHA is shown in Fig. 4. The most important 2 steps is:

Step 1 (Calculate the transition probability matrix of DDS, MMS and MD): The transition probability matrix $M = (M(i, j))_{n_m \times n_m}$ of MMS is constructed as:

$$M(i, j) = \begin{cases} \frac{M_S(i,j)}{\sum_{k=1}^{n_m} M_S(k,j)}, & \sum_{k=1}^{n_m} M_S(k, j) \neq 0 \\ 0, & \text{otherwise} \end{cases}$$

(12)

Similarly, $D(D(i, j))_{n_d \times n_d}$ is the transition probability matrix of the DDS:

$$D(i, j) = \begin{cases} \frac{D_S(i,j)}{\sum_{k=1}^{n_d} D_S(k,j)}, & \sum_{k=1}^{n_d} D_S(k, j) \neq 0 \\ 0, & \text{otherwise} \end{cases}$$

(13)

Based on hybrid recommendation algorithm, the miRNA node m_i is assigned an initial lever of resource $f(m_i) = 1$, *or* 0 depending on whether the miRNA is associated with given disease. All the resource of miRNA nodes redistributed via the transition matrix of hybrid recommendation algorithm, and transition probability matrix of MD is calculated as:

$$P_A = W^{H+P} * A.$$

(14)

Step 2 (Implement unbalance bi-random walk in heterogeneous network): Because of the different topological characteristics between MMS and DDS, in these two networks, we introduce two parameters of l and r as the biggest step random walk on MMS and DDS.
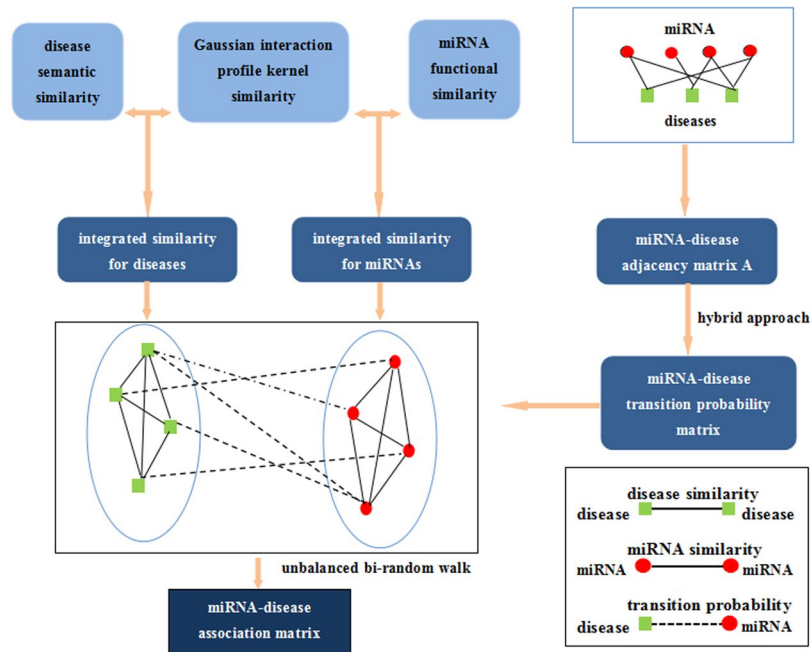
Figure 4. Flowchart of potential miRNA-disease association prediction based on the computational model of BRWHNHA.

$$MMS: P_{t\text{-}M} = (1 - \alpha) * M * P_{t-1} + \alpha * P_A \tag{15}$$

$$DDS: P_{t\text{-}D} = (1 - \alpha) * P_{t-1} * D + \alpha * P_A \tag{16}$$

$$P_t = \begin{cases} \frac{P_{t_M} + P_{t_D}}{2} & t \leq r, t \leq l \\ P_{t\text{-}M}, & t \leq r, t > l \\ P_{t\text{-}D}, & t > r, t \leq l \end{cases} \tag{17}$$

α denotes a decay factor ranging from 0 and 1. The matrix P_A is used to control the prior probability of the iterative process and is the transition probability matrix of the bipartite network G obtained by the recommendation algorithm. P_A is a transition probability matrix, and $P_0 = P_A/sum(P_A)$. After several iterations, P_t is the steady-state probability matrix between miRNAs and diseases. For a given disease, we ranked all the candidate miRNAs based on the probability. In BRWHNHA algorithm, we effectively utilize the topological information of heterogeneous networks, including: MMS, DDS and MD.

Data Availability
All data generated or analyzed during this study are included in this article [Additional file 2, Additional file 3, Additional file 4]. (The data we used was downloaded from the paper of Luo and Xiao[22]).

References
1. Lee, R. C., Feinbaum, R. L. & Ambros, V. The c. elegans heterochronic gene lin-4 encodes small rnas with antisense complementarity to lin-14. *Cell* **75**, 843–854 (1993).
2. Wang, C., Wei, L., Guo, M. & Quan, Z. Computational approaches in detecting non-coding rna. *Current Genomics* **14**, 371 (2013).
3. Wei, L. *et al*. Improved and promising identification of human micrornas by incorporating a high-quality negative set. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **11**, 192–201 (2014).
4. Mitra *et al*. Identifying transcription factor and microrna mediated synergetic regulatory networks in lung cancer. *BMC Bioinformatics* **14**, A14 (2013).
5. Cheng, A. M., Byrom, M. W., Shelton, J. & 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).
6. Miska, E. How micrornas control cell division, differentiation and death. *Current Opinion Genetics Development* **15**, 563–568 (2005).
7. Xu, P., Guo, M. & Hay, B. A. Micrornas and the regulation of cell death. *TRENDS Genetics* **20**, 617–624 (2004).
8. Wu, D. *et al*. ncrdeathdb: A comprehensive bioinformatics resource for deciphering network organization of the ncrnamediated cell death system. *Autophagy* **11**, 1917–1926 (2015).
9. Li, Y., W., Y. & Zhuang, L. Connect the dots: a systems level approach for analyzing the mirnar -mediated cell death network. *Autophagy* **9**, 436–439 (2013).
10. Kahraman, M. *et al*. Microrna in diagnosis and therapy monitoring of early-stage triple-negative breast cancer. *Scientific Reports* **8**, 11584 (2018).

11. Markou, A. *et al*. Prognostic value of mature microrna-21 and microrna-205 overexpression in non-small cell lung cancer by quantitative real-time rt-pcr. *Clinical Chemistry* **54**, 1696–1704 (2008).
12. Miller, TylerE. *et al*. Microrna-221/222 confers tamoxifen resistance in breast cancer by targeting p27kip1. *Journal of Biological Chemistry* **283**, 29897–29903 (2008).
13. Weinberg, M. S. & Wood, M. J. A. Short non-coding rna biology and neurodegenerative disorders. novel disease targets and therapeutics. *Human Molecular Genetics* **18**, R27–R39 (2009).
14. Jiang, Q. *et al*. Prioritization of disease micrornas through a human phenome-micrornaome network. *BMC Systems Biology* **4**, 1–9 (2010).
15. Jiang, Q., Hao, Y., Wang, G., Zhang, T. & Wang, Y. Weighted networkbased inference of human microrna-disease associations. *Fifth International Conference on Frontier of Computer Science and Technology* August **18**–22, 431–435 (2010).
16. Chen, X., Liu, M. X. & Yan, G. Y. Rwrmda: Predicting novel human microrna-disease associations. *Molecular Biosystems* **8**, 2792–2798 (2012).
17. Xuan, P. *et al*. Prediction of potential disease-associated micrornas based on random walk. *Bioinformatics* **31**, 1805–1815 (2015).
18. Chen, X. *et al*. Wbsmda: Within and between score for mirna-disease association prediction. *Scientific Reports* **6**, 21106 (2016).
19. 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 (2016).
20. Liu, Y., Zeng, X., He, Z. & Zou, Q. Inferring microrna-disease associations by random walk on a heterogeneous network with multiple data sources. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* (2017).
21. Chen, X., Niu, Y. W., Wang, G. H. & Yan, G. Y. Hamda: hybrid approach for mirna-disease association prediction. *Journal of Biomedical Informatics* **76**, 50–58 (2017).
22. Luo, J & Qiu, X. A novel approach for predicting microrna-disease associations by unbalanced bi-random walk on heterogeneous network. *Journal of Biomedical Informatics* **66**, 194–203 (2017).
23. Zeng, X., Liu, L., Lv, L. & Zou, Q. Prediction of potential disease-associated micrornas using structural perturbation method. *Bioinformatics* **PP**, 1–1 (2018).
24. Yang, Z. *et al*. Dbdemc: a database of (2010) differentially expressed mirnas in human cancers. *BMC Genomics* **11**, 1–8 (2010).
25. Jiang *et al*. mir2disease: a manually curated database for microrna deregulation in human disease. *Nucleic Acids Research* **37**(Database issue), D98–104 (2008).
26. McGuire, S. World cancer report 2014. geneva, switzerland: World health organization, international agency for research on cancer, who press, 2015. *Advances in Nutrition* **7**, 418–419 (2016).
27. Hart, M. *et al*. The protooncogene erg is a target of microrna mir-145 in prostate cancer. *Febs Journal* **280**, 2105–2116 (2013).
28. Ueno, K. *et al*. Microrna-183 is an oncogene targeting dkk-3 and smad4 in prostate cancer. *British Journal of Cancer* **108**, 1659–1667 (2013).
29. Lu, M. *et al*. An analysis of human microrna and disease associations. *PLoS One* **3**, e3420 (2008).
30. Bandyopadhyay, S., Mitra, R., Maulik, U. & Zhang, M. Q. Development of the human cancer microrna network. *Silence* **1**, 6 (2010).
31. Wang, J. Z. *et al*. A new method to measure the semantic similarity of go terms. *Bioinformatics* **23**, 1274–1281 (2007).
32. Chen, X. & Yan, G.-Y. Novel human lncrna-disease association inference based on lncrna expression profiles. *Bioinformatics* **29**, 2617–2624 (2013).
33. Chen, X. *et al*. Nllss: Predicting synergistic drug combinations based on semi-supervised learning. *PLoS Computational Biology* **12**, e1004975 (2016).
34. Zhou, L., Liu, K., Liu, J. & Zhang, R. Solving the apparent diversity-accuracy dilemma of recommender systems. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 4511–4515 (2010).

Acknowledgements

Author Contributions

D.L.Y. and Y.L.M. contributed to the conception and design of the study and developed the method. D.L.Y. implemented the algorithms and analyzed the data and results. Z.G.Y. gave the ideas and supervised the project. D.L.Y. wrote the manuscript. All authors discussed the results and reviewed the manuscript, and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-019-39226-x.

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.