SCIENTIFIC REP🞓RTS

OPEN

miRDDCR: a miRNA-based method to comprehensively infer drug-disease causal relationships

Hailin Chen[1], Zuping Zhang[2] & Wei Peng[3]

Revealing the cause-and-effect mechanism behind drug-disease relationships remains a challenging task. Recent studies suggested that drugs can target microRNAs (miRNAs) and alter their expression levels. In the meanwhile, the inappropriate expression of miRNAs will lead to various diseases. Therefore, targeting specific miRNAs by small-molecule drugs to modulate their activities provides a promising approach to human disease treatment. However, few studies attempt to discover drug-disease causal relationships through the molecular level of miRNAs. Here, we developed a miRNA-based inference method miRDDCR to comprehensively predict drug-disease causal relationships. We first constructed a three-layer drug-miRNA-disease heterogeneous network by combining similarity measurements, existing drug-miRNA associations and miRNA-disease associations. Then, we extended the algorithm of Random Walk to the three-layer heterogeneous network and ranked the potential indications for drugs. Leave-one-out cross-validations and case studies demonstrated that our method miRDDCR can achieve excellent prediction power. Compared with related methods, our causality discovery-based algorithm showed superior prediction ability and highlighted the molecular basis miRNAs, which can be used to assist in the experimental design for drug development and disease treatment. Finally, comprehensively inferred drug-disease causal relationships were released for further studies.

Most drugs achieve their therapeutic functions by binding to specific molecular targets which are relevant to an abnormal state, thereby changing the biochemical and/or biophysical activities of these molecules[1]. Therefore, it is of critical importance to investigate how drugs and diseases form their causal relationships in the molecular level. Many evidences suggest that a drug can act on multiple targets rather than one target[2,3]. More recently, increasing studies revealed that drugs can target microRNAs (miRNAs), which are short (~22 nucleotide) non-coding RNAs, and regulate their expressions. For instance, Rossi *et al.*[4] detected that the expression levels of 22 miRNAs were altered with the treatment of 5-fluorouracil in human colon cell line.

miRNAs are single-stranded RNAs with post transcriptional regulatory functions. They regulate gene expressions by base pairing to complementary sequences of their target mRNAs[5–7]. Thus, accumulating researches indicated that miRNAs are involved in a broad range of biological processes, such as cellular signaling[8], proliferation[9,10] and metabolism[11]. As such, changes in the expression levels of particular miRNAs are related to many kinds of critical diseases. For example, the reduced expression level of let-7 was shown to be associated with lung cancer progression[12]. In addition, miRNAs are suitable to be drug targets as they have several attractive features, such as specific secondary structures and conserved sequences[13]. As a result, restoring miRNA expression levels with small-molecule drugs offers an innovative and promising approach to human disease treatment[13], which has brought forward to a new research field of miRNAs in pharmacogenomics[14].

With this understanding, studying drug-disease causal relationships under the perspective of their genetic basis miRNAs could not only help to unveil mechanisms of action of drugs but also advance public health. However, few researches address this question in a systematic view.

Based on chemical descriptors and machine learning algorithms, Jamal *et al.*[15] created computational models to predict drugs' biological activities on miRNAs. Jiang *et al.*[16] applied transcriptional responses to identify associations between drugs and cancer-related miRNAs. Depending on gene expression signatures of bioactive

[1]School of Software, East China Jiaotong University, Nanchang, China. [2]School of Information Science and Engineering, Central South University, Changsha, China. [3]Computer Center of Kunming University of Science and Technology, Kunming, China. Correspondence and requests for materials should be addressed to H.C. (email: chl_ecjtu@hotmail.com)
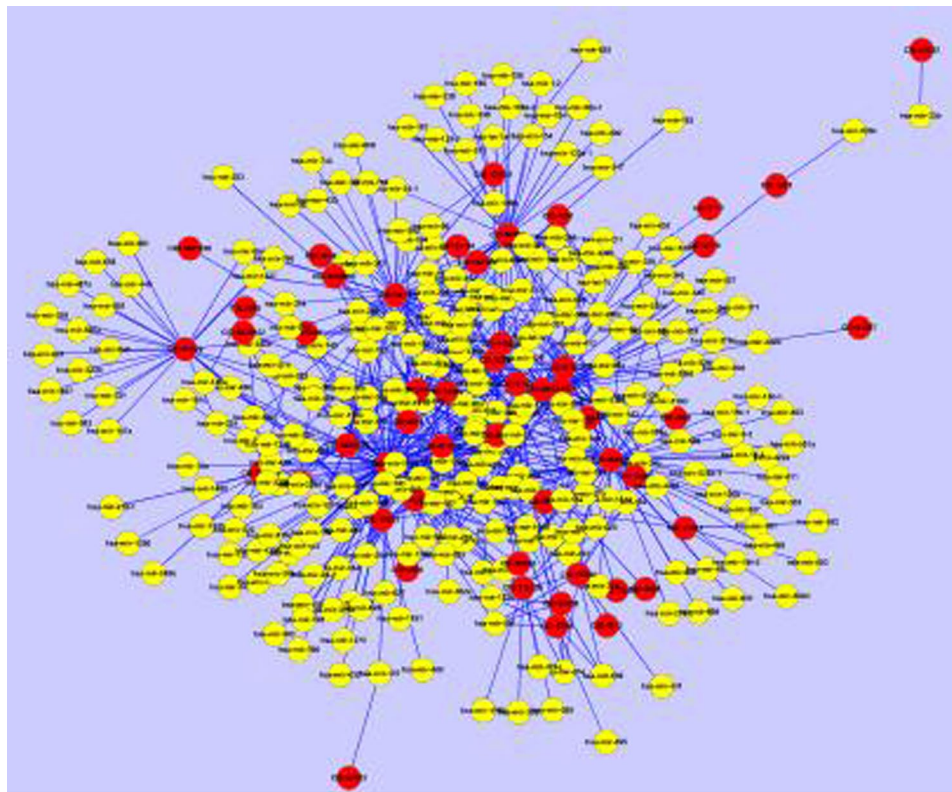
Figure 1. The drug-miRNA bipartite graph. The red circles represent drugs and the pink circles denote miRNAs. This graph was prepared by using the 630 experimentally confirmed drug-miRNA associations.

small molecule perturbation and Alzheimer's disease (AD)-related miRNA regulation, Meng *et al.*[17] presented a systematic computational approach to constructing a drug-miRNA association network in AD. By integrating features of drugs and miRNAs, Lv *et al.*[18] proposed an algorithm of Random Walk with Restart (RWR) to infer new associations between small molecules and miRNAs. The above four methods all identify only drug-miRNA associations, but do not comprehensively provide therapeutic potential for drugs.

In the meantime, as miRNAs are highly relevant to multiple complex diseases, many computational efforts[19–33] have been devoted to detecting potential miRNA-disease associations for understanding the molecular mechanisms of diseases. These methods mainly relay on the assumption that miRNAs tend to show similar dysfunctional evidences for similar disease clusters[34]. Generally, for these researches various features were first integrated for miRN-miRNA and disease-disease similarity calculation. Network-based or machine learning-based algorithms were then developed to rank the most promising disease-related miRNAs or miRNA-related diseass for further biomedical tests. These studies provide reliable guidance for *in vivo* experiment design. However, one limitation lied in these methods is that they could not provide information of disease treatment.

For the above studies, predictions of drug-miRNA associations and miRNA-disease associations were treated separately. We argue that by incorporating information of target miRNAs, we can make more insightful drug-disease causal relationship predictions. Based on the assumption that drugs will form relationships with diseases when they share some significant miRNA partners, Chen *et al.*[35] applied hyper-geometric tests by combining existing drug-miRNA associations and miRNA-disease associations to predict drug-disease associations. This is the first computational model proposed to infer drug-disease associations, in which the molecular basis miRNAs is explicitly included. Even though a high AUC value could be received, this method is not workable for new drugs whose drug-miRNA associations cannot be obtained.

In this paper, we developed a miRNA-based method to extensively predict drug-disease causal relationships (miRDDCR). Based on two previous researches[18,34], the proposed method miRDDCR relied on the hypothesis that similar small molecules tend to target similar miRNAs, and finally treat similar diseases. miRDDCR can predict drug-disease relationships in a large scale by combining similarity measurements, existing drug-miRNA associations and miRNA-disease associations. To evaluate the prediction performance of our method, leave-one-out cross validations (LOOCV) were conducted and satisfied AUC values could be received. Compared with existing methods, our method miRDDCR shows superior prediction ability. Moreover, case studies of two drugs demonstrated that our method is powerful in predicting drug-disease causal relationships with a high level of reliability. After validating the usefulness of our method, we used miRDDCR to comprehensively infer drug-disease causal relationships, which we hope will facilitate further drug discovery and disease treatment.
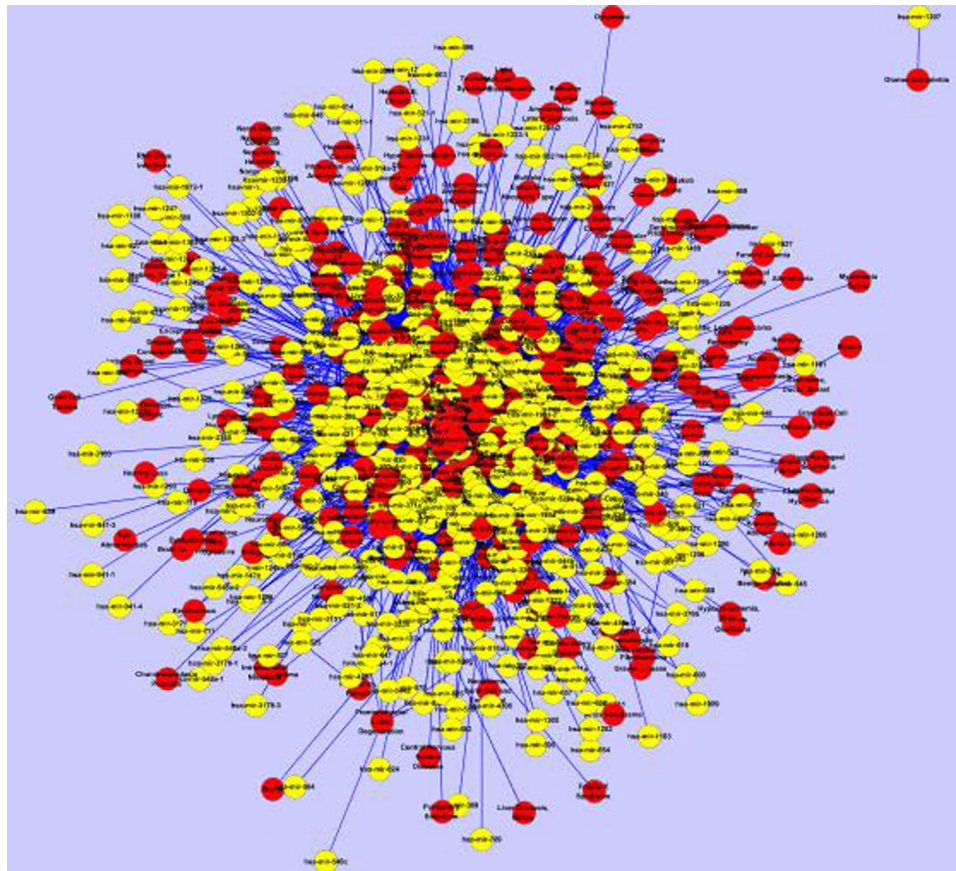
Figure 2. The miRNA-disease bipartite graph. The red circles indicate diseases and the pink circles denote miRNAs. This graph was drawn by using the 6082 known miRNA-disease associations.

No. of drugs	No. of miRNAs	No. of drug-miRNA associations	Average degree of drugs	Average degree of miRNAs	Sparsity
831	540	630	0.76	1.17	0.0014

Table 1. Statistics of the drug-miRNA bipartite graph.

No. of miRNAs	No. of diseases	No. of miRNA-disease associations	Average degree of miRNAs	Average degree of diseases	Sparsity
540	341	6082	11.26	17.84	0.033

Table 2. Statistics of the miRNA-disease bipartite graph.

Results

Preliminary analysis of datasets used in this manuscript. The datasets (Supplementary Dataset S1) used in our paper consisted of 831 drugs, 540 miRNAs, 341 diseases, 630 drug-miRNA associations and 6082 miRNA-disease associations. A whole view of the drug-miRNA bipartite graph and the miRNA-disease bipartite graph could be seen in Figs 1 and 2. For the 831 drugs, there were only 51 drugs whose drug-miRNA associations exist. Some statistical analysis of the two bipartite networks was listed at Tables 1 and 2, respectively. We could observe that both the two bipartite graphs were sparse.

Parameter tuning and performance evaluation of the proposed method miRDDCR. There exist 6 parameters involved in our algorithm. The parameters α_1 and α_2 were decay factors. The other 4 parameters l_1, r_1, l_2 and r_2 were considered as the numbers of maximal iterations of random walks on the bipartite networks. For parameter tuning, we followed ref.[36] to set $\alpha_1 = \alpha_2 = 0.8$ and $l_1 = r_1 = l_2 = r_2 = 4$.

The predicted indication results for the whole 831 drugs were ranked according to the final values received from the algorithm miRDDCR. A bigger value indicated a greater probability that a drug forms a causal relationship with a disease. Experimentally validated drug-disease relationships were extracted from Comparative Toxicogenomics Database (CTD)[37], DrugBank[38] and Therapeutic Targets Database (TTD)[39]. We collected 13490
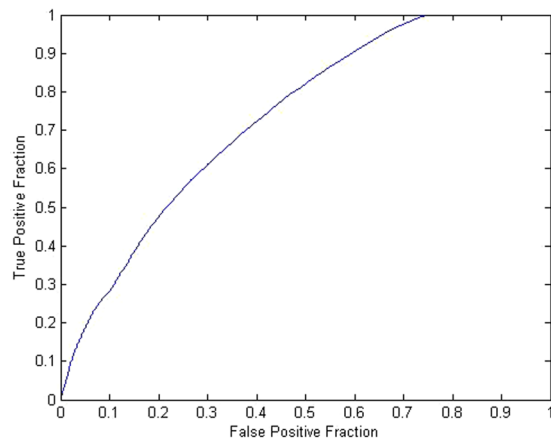
Figure 3. Performance evaluation of miRDDCR in term of ROC curve.
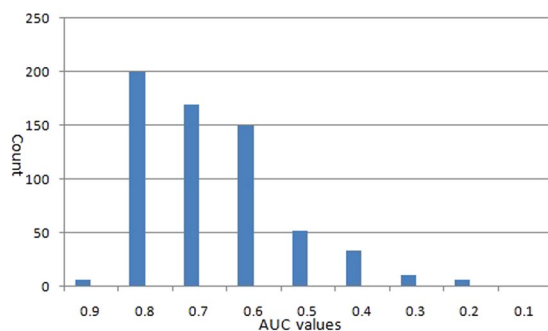

Figure 4. The distribution of AUC values received by leave-one-drug-out cross-validations (LOOCV) for the 630 drugs.

drug-disease relationships which were relevant to our study from the three databases. These confirmed relationships were used as a gold standard dataset for performance evaluation. Taking the known 13490 relationships as the positive instances, a receiver operating characteristics (ROC) curve (see Fig. 3) was drawn by calculating true positive fraction (TPR, sensitivity) and false positive fraction (FPR, 1-specificity) at different cutoffs. Finally, a value of area under curve (AUC) of 0.7334 was received.

Furthermore, we used leave-one-drug-out cross-validations (LOOCV) to evaluate the performance of miRD-DCR in predicting drug-disease causal relationships. For the whole 831 drugs, each drug was considered as a test drug once and all its drug-miRNA association information was removed. The remaining 830 drugs were taken as the training dataset. For the 831 drugs, there existed 201 drugs whose experimentally confirmed drug-disease relationships were not available in the gold standard dataset. Therefore, we could not calculate their AUC values. For the remaining 630 drugs, the distribution of their AUC values, when leave-one-drug-out cross validations were implemented, could be seen in Fig. 4 with an average AUC value of 0.717. Meanwhile, for about 60% (376/630) of the 630 drugs, we received higher AUC values than 0.7, with the highest AUC value of 0.985 for the drug CID:3108 (*Dipyridamole*).

Comparison with existing methods. Until recently, efforts made on causality discovery-based methods for drug-disease relationship predictions were rare. The most related study to ours is the inference model introduced in ref.[35], in which drug-disease causality relationship predictions were only based on known drug-miRNA associations and miRNA-disease associations. For novel drugs, whose drug-miRNA associations were not available, their indications could not be inferred by this method. Our method miRDDCR overcame this drawback by taking advantage of similarity measurements. In the leave-one-drug-out cross-validation section, we considered each drug as a novel drug and reliable prediction ability could be obtained.

With the accumulation of biochemical data, several algorithms[40–43] have been put forward to predict potential drug–disease relationships. They first integrated multiple types of features to construct drug-drug and disease-disease similarity metrics. Depending on the similarity values and experimentally verified drug-disease relationships, machine learning-based or network-based methods were then developed to predict new indications for drugs. Different sources of information prevented a direct comparison between these methods and miRDDCR. Generally, these methods relayed heavily on known drug-disease relationships for new drug-disease association prediction. Obviously, experimentally confirmed drug-disease associations are scarce. The sparsity
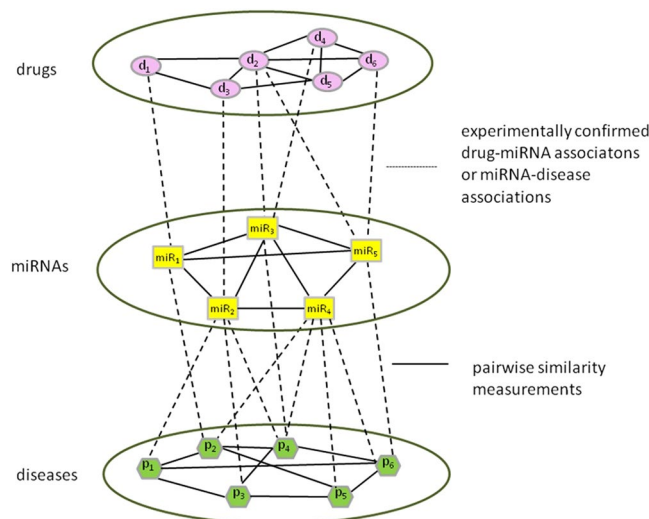
Figure 5. A three-layer drug–miRNA-disease heterogeneous network.

of data preparation might influence prediction accuracy. Our method did not need known drug-disease associations for prediction. Moreover, these methods, unlike our method miRDDCR, provided no molecular hints to help design experiments to test and confirm the results. Our method made full use of the information of target miRNAs and provided valuable resource for drug design and disease treatment.

Case studies.   In this experimental scenario, case studies of the two drugs, CID:33887 (*Almitrine*) and CID:137 (*Aminolevulinic acid*), were analyzed for further evaluation of the ability of our method miRDDCR to predict potential drug-disease causal relationships.

For the drug CID:33887 (*Almitrine*), there was no target miRNA information in the prepared drug-miRNA dataset. After the first round of bi-random walks in miRDDCR on the drug-miRNA bipartite network, potential target miRNAs were ranked, in which the top 5 miRNAs were hsa-mir-21, hsa-mir-27b, hsa-mir-23a, hsa-mir-27a and hsa-mir-155. With the second round of bi-random walks on the miRNA-disease bipartite network, the indications of the drug *Almitrine* could be received. The top 10 predicted diseases were breast neoplasms, hepatocellular carcinoma, stomach neoplasms, colorectal neoplasms, melanoma, lung neoplasms, neoplasms, ovarian neoplasms, heart failure and prostatic neoplasms. The 9th (9/341) disease heart failure was the only indication available for the drug in the gold standard dataset and we successfully predicted this relationship with a high rank. It should be noted that 4 (hsa-mir-21, hsa-mir-23a, hsa-mir-27a and hsa-mir-155) out of the top 5 target miRNAs, when expressed abnormally, were involved in the development of heart failure[44].

For the drug CID:137 (*Aminolevulinic acid*), its potential indications were ranked according to the scores received by our algorithm miRDDCR. We selected the top 5, top 10, top 20, top 30 and top 40 predicted diseases and found that there were 5, 9, 17, 21 and 23 results supported by the gold standard dataset, respectively. We took the top 1 predicted disease breast neoplasms as an example to explain the inference ability of our algorithm. For the drug CID:137 (*Aminolevulinic acid*), there was no target miRNA information in the 630 drug-miRNA associations. After the first round of random walk, we chose the top 10 predicted target miRNAs and discovered that the 3rd target (hsa-mir-450a-2), the 5th target (hsa-mir-23a), the 6th target (hsa-mir-29b-2), the 7th target (hsa-mir-320b-1) and the 10th target (hsa-mir-375) were associated with the predicted disease breast neoplasms[44].

As currently confirmed drug-disease associations were not complete, we think the other predicted diseases with high ranks could be potential indications for the drugs.

Comprehensive drug-disease causal relationship predictions.   After verifying the prediction power of our method by cross-validations and case studies, all the known associations were used as training data to comprehensively predict potential drug-disease causal relationships. Moreover, as causal factors in the molecular level, the top 20 predicted target miRNAs for each drug (Supplementary Dataset S2 online) was available for further experiment tests. For each of the 831 drugs, we published the top 50 predicted candidate indications for future studies. The full list of the whole inferred relationships can be obtained from the Supplementary Dataset S3 online. To be more accurate, we suggested taking the prediction results for the 376 drugs, whose AUC values were greater than 0.7 in leave-one-out cross validations, into consideration.

Discussion

Exploring the molecular mechanism behind the curative effects of drugs is crucial for drug development and disease treatment. Recent researches suggested that drugs can achieve their therapeutic functions by targeting miRNAs because the abnormal expression of miRNAs could lead to a lot of complex diseases and drugs can bind miRNAs to restore their expression levels. As a relatively new discipline in biomedical research, our understanding of drug-targeted miRNAs and miRNA-related diseases is lacking. Furthermore, few studies attempted to systematically discover drug-disease causal relationships through the target miRNAs.

$\text{miRDDCR}(S_d, S_{miR}, S_p, E_1, E_2, \alpha_1, \alpha_2, l_1, r_1, l_2, r_2)$

// α_1 and α_2 were decay factors. l_1, r_1, l_2 and r_2 were considered as the numbers of maximal

//iterations of random walks on the left and right networks

1. normalize S_d, S_{miR}, S_p

2. $R_0 = E_1 / sum(E_1), T_0 = E_2 / sum(E_2)$

3. for $s1 = 1$ to $\max(l_1, r_1)$ // random walk on two-layer drug-miRNA network

4.   if $s1 \leq l_1$

5. $R_{s1\_left} = \alpha_1 \times normalized(S_d) \times R_{s1-1} + (1-\alpha_1) \times E_1$

6.   if $s1 \leq r_1$

7. $R_{s1\_right} = \alpha_1 \times R_{s1-1} \times normalized(S_{miR}) + (1-\alpha_1) \times E_1$

8. $R_{s1} = (\delta_{s_1 \leq l_1} \times R_{s1-left} + \delta_{s_1 \leq r_1} \times R_{s1-right}) / (\delta_{s_1 \leq l_1} + \delta_{s_1 \leq r_1})$ // $\delta_{x \leq y}$ is 1 if $x \leq y$ and

//0 otherwise

9. end for

10. for $s2 = 1$ to $\max(l_2, r_2)$ // random walk on two-layer miRNA-disease network

11.   if $s2 \leq l_2$

12. $T_{s2\_left} = \alpha_2 \times normalized(S_{miR}) \times T_{s2-1} + (1-\alpha_2) \times E_2$

13.   if $s2 \leq r_2$

14. $T_{s2\_right} = \alpha_2 \times T_{s2-1} \times normalized(S_p) + (1-\alpha_2) \times E_2$

15. $T_{s2} = (\delta_{s_2 \leq l_2} \times T_{s2-left} + \delta_{s_2 \leq r_2} \times T_{s2-right}) / (\delta_{s_2 \leq l_2} + \delta_{s_2 \leq r_2})$ // $\delta_{x \leq y}$ is 1 if $x \leq y$

//and 0 otherwise

16. end for

17. return $R \times T$

Figure 6. Description of workflows of the algorithm miRDDCR.

In this paper, we developed an approach miRDDCR to revealing the causal relationships between drugs and diseases under their molecular basis miRNAs. Our method makes full use of similarity measurements, known drug-miRNA and miRNA-disease associations to infer indications for drugs. As current drug-miRNA and miRNA-disease associations are insufficient, two rounds of bi-random walks are implemented to reveal the hidden associations in the drug-miRNA-disease heterogeneous network. We have applied our method to real datasets and the results showed that our method could successfully discover the causality underlying drugs and diseases. Compared with other algorithms, our method provided a straightforward strategy for causality discovery and showed superior prediction performance.

As similarity values provide a vital role in the prediction procedure, we need to further investigate the types of features collected for similarity measurements. Meanwhile, the predicted drug-disease relationships might be biased as the numbers of known drug-miRNA associations and miRNA-disease associations were rare. We expect the prediction power of our method can be improved by integrating more experimentally confirmed drug-miRNA and miRNA-disease associations.

It should be noted that the mechanism of action of drugs has not been completely investigated. The most recent studies indicated that drugs may also target other non-coding RNAs, including circular RNAs, long non-coding RNAs and Piwi-interacting RNAs (see ref.[45] for more details). At the same time, complex diseases are multi-factor driven. For example, the study conducted by Wang et al.[46] suggested that DNA mutations and other genomic alterations, which were valuable biomarkers and could be used to construct disease hallmark networks, provided significant roles in cancer clonal evolution and associated clinical phenotypes. Therefore, more relevant biochemical information is needed for deepening our knowledge of drug-disease causal relationships, even though our study provided a feasible strategy for discovering the causality.

Methods

Datasets. The 831 drugs used in our manuscript were downloaded from ref.[18]. Similar to ref.[18], the pairwise similarity values of those drugs were calculated by integrating information of chemical structures[47], functional consistency[48] and side effects[40]. The integrated similarity S_d was defined as:

$$S_{\mathrm{d}} = (S_d^c + S_d^f + S_d^s)/3 \qquad\qquad (1)$$

Here, S_d^c, S_d^f and S_d^s denoted the similarity measurements based on chemical structures, functional consistency and side effects, respectively.

The 540 miRNAs employed in this paper were also collected from ref.[18]. Similar to ref.[18], the pairwise similarity S_{miR} of those miRNAs was based on functional consistency[48].

We obtained 341 diseases and their pairwise similarity measurements S_{p} from ref.[26]. The similarity S_{p} was calculated by incorporating disease semantic similarity and disease functional similarity.

Experimentally validated drug-miRNA associations and miRNA-disease associations were selected from the latest versions of SM2miR[49] and HMDD[44], respectively. For drug-miRNA association retrieval, we restricted the species to Homo sapiens. After removing duplicate records stored in the databases of SM2MiR and HMDD, we finally received 630 drug-miRNA associations and 6082 miRNA-disease associations.

Method Description. We defined the drugs, miRNAs and diseases as $D = \{d_1, d_2, \ldots, d_i\}$, $miR = \{miR_1, miR_2, \ldots, miR_j\}$ and $P = \{p_1, p_2, \ldots, p_k\}$. Experimentally confirmed drug-miRNA associations were modeled as a bipartite graph $G1 = \{V1, E1\}$, where $V1 = \{D, miR\}$ and $E1 = \{a_{ij} : d_i \in D, miR_j \in miR\}$. For $E1$, its values were 1 or 0 which indicated the presence or absence of each association. Similarly, existing miRNA-disease associations were considered as another bipartite graph $G2 = \{V2, E2\}$, where $V2 = \{miR, P\}$ and $E2 = \{r_{mn} : miR_m \in miR, p_n \in P\}$. For $E2$, its values were also 1 or 0 which represented the presence or absence of each association.

We connected the drug-miRNA bipartite graph and the miRNA-disease bipartite graph together to construct a three-layer drug-miRNA-disease heterogeneous network. For each layer, an edge was drawn between two nodes when the similarity of the two nodes was bigger than 0. The weight of the edge was set to be the similarity value. An example of the heterogeneous network was illustrated in Fig. 5. The objective of this research is to predict drug-disease causal relationships based on the three-layer heterogeneous network.

Previously, random walk[36,50–53] has been widely used in bipartite graphs for bilateral association prediction in bioinformatics. For example, bi-random walk (BiRW)[36] was successfully applied on both gene network and phenotype network simultaneously, with an averaged output from the two networks in each step, to infer potential gene-phenotype associations. Inspired by the successful application of BiRW, we extended the algorithm to the three-layer drug–miRNA-disease heterogeneous network to predict potential drug-disease causal relationships.

In this study, a three-layer heterogeneous network, including a layer of causal factors of miRNAs, was constructed and the drug-disease causal relationship prediction process mainly included three steps. First, bi-random walks were applied on the two-layer drug-miRNA network. Second, another round of bi-random walks was established on the two-layer miRNA-disease network. Finally, the prediction results were received by combining the outcomes from the two previous steps. The whole pipeline of the algorithm miRDDCR could be outlined in Fig. 6.

References

1. Hopkins, A. L. & Groom, C. R. The druggable genome. *Nat. Rev. Drug Discovery.* **1**, 727–730 (2002).
2. Anighoro, A., Bajorath, J. & Rastelli, G. Polypharmacology: challenges and opportunities in drug discovery: miniperspective. *J. Med. Chem.* **57**, 7874–7887 (2014).
3. Yıldırım, M. A., Goh, K. I., Cusick, M. E., Barabási, A. L. & Vidal, M. Drug-target network. *Nat. Biotechnol.* **25**, 1119–1126 (2007).
4. Rossi, L., Bonmassar, E. & Faraoni, I. Modification of miR gene expression pattern in human colon cancer cells following exposure to 5-fluorouracil *in vitro*. *Pharmacol. Res.* **56**, 248–253 (2007).
5. Ambros, V. The functions of animal microRNAs. *Nature* **431**, 350–355 (2004).
6. Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–297 (2004).
7. Bartel, D. P. MicroRNAs: target recognition and regulatory functions. *Cell* **136**, 215–233 (2009).
8. Cui, Q., Yu, Z., Purisima, E. O. & Wang, E. Principles of microRNA regulation of a human cellular signaling network. *Mol. Syst. Biol.* **2**, 46 (2006).
9. Chen, J. F. *et al.* The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. *Nat. Genet.* **38**, 228–233 (2006).
10. Zhao, Y., Samal, E. & Srivastava, D. Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis. *Nature* **436**, 214–220 (2005).
11. Poy, M. N. *et al.* A pancreatic islet-specific microRNA regulates insulin secretion. *Nature* **432**, 226–230 (2004).
12. Esquela-Kerscher, A. *et al.* The let-7 microRNA reduces tumor growth in mouse models of lung cancer. *Cell Cycle* **7**, 759–764 (2008).
13. Zhang, S., Chen, L., Jung, E. J. & Calin, G. A. Targeting microRNAs with small molecules: from dream to reality. *Clin. Pharmacol. Ther.* **87**, 754–758 (2010).
14. Shomron, N. MicroRNAs and pharmacogenomics. *Pharmacogenomics* **11**, 629–632 (2010).
15. Jamal, S. *et al.* Computational analysis and predictive modeling of small molecule modulators of microRNA. *J. Cheminformatics* **4**, 16 (2012).
16. Jiang, W. *et al.* Identification of links between small molecules and miRNAs in human cancers based on transcriptional responses. *Sci. Rep.* **2**, 282 (2012).
17. Meng, F. *et al.* Constructing and characterizing a bioactive small molecule and microRNA association network for Alzheimer's disease. *J. R. Soc. Interface* **11**, 20131057 (2014).
18. Lv, Y. *et al.* Identifying novel associations between small molecules and miRNAs based on integrated molecular networks. *Bioinformatics* **31**, 3638–3644 (2015).
19. Chen, H. & Zhang, Z. Prediction of associations between OMIM diseases and MicroRNAs by random walk on OMIM disease similarity network. *The Scientific World J.* **2013**, 204658 (2013).
20. Chen, H. & Zhang, Z. Similarity-based methods for potential human microRNA-disease association prediction. *BMC Med. Genomics* **6**, 12 (2013).
21. Chen, X. & Yan, G.-Y. Semi-supervised learning for potential human microRNA-disease associations inference. *Sci. Rep.* **4**, 5501 (2014).

22. Zou, Q. *et al.* Prediction of microRNA-disease associations based on social network analysis methods. *BioMed Res. Int.* **2015**, 810514 (2015).

23. Liao, B., Ding, S., Chen, H., Li, Z. & Cai, L. Identifying human microRNA–disease associations by a new diffusion-based method. *J. Bioinf. Comput. Boil.* **13**, 1550014 (2015).

24. Lan, W. *et al.* Predicting microRNA-disease associations based on improved microRNA and disease similarities. *IEEE ACM T. Comput. Bi.* **2016**, PP (99):1–1 (2016).

25. Luo, J., Ding, P., Liang, C., Cao, B., & Chen, X. Collective prediction of disease-associated miRNAs based on transduction learning. *IEEE ACM T. Comput. Bi.* **2016**, PP (99):1–1 (2016).

26. 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 T. Comput. Bi.* **2016**, PP (99):1–1 (2016).

27. Chen, X. *et al.* HGIMDA: Heterogeneous graph inference for miRNA-disease association prediction. *Oncotarget* **7**, 65257–65269 (2016).

28. Chen, X. *et al.* WBSMDA: within and between score for MiRNA-disease association prediction. *Sci. Rep.* **6**, 21106 (2016).

29. Peng, W., Lan, W., Yu, Z. *et al.* A Framework for integrating multiple biological networks to predict microRNA-disease associations. *IEEE T. Nanobiosci.* **16**, 100–107 (2017).

30. Yu, H., Chen, X. & Lu, L. Large-scale prediction of microRNA-disease associations by combinatorial prioritization algorithm. *Sci. Rep.* **7**, 43792 (2017).

31. You, Z. H. *et al.* PBMDA: A novel and effective path-based computational model for miRNA-disease association prediction. *PLoS Comput. Biol.* **13**, e1005455 (2017).

32. Chen, X., Wu, Q. F. & Yan, G. Y. RKNNMDA: Ranking-based KNN for MiRNA-Disease Association prediction. *RNA Biol.* **2017**, 1–11 (2017).

33. Chen, X., Gong, Y., Zhang, D. H., You, Z. H. & Li, Z. W. DRMDA: deep representations-based miRNA–disease association prediction. *J. Cell. Mol. Med.* **2017** (Suppl. 1) (2017).

34. Lu, M. *et al.* An analysis of human microRNA and disease associations. *PloS one* **3**, e3420 (2008).

35. Chen, H. & Zhang, Z. A miRNA-driven inference model to construct potential drug-disease associations for drug repositioning. *BioMed Res. Int.* **2015**, 406463 (2015).

36. Xie, M., Hwang, T. & Kuang, R. Prioritizing disease genes by bi-random walk. *Advances in Knowledge Discovery and Data Mining* **2012**, 292–303 (2012).

37. Davis, A. P. *et al.* The comparative toxicogenomics database: update 2013. *Nucleic Acids Res.* **41**, D1104–D1114 (2012).

38. Law, V. *et al.* DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res.* **42**, D1091–D1097 (2013).

39. Yang, H. *et al.* Therapeutic target database update 2016: enriched resource for bench to clinical drug target and targeted pathway information. *Nucleic Acids Res.* **44**, D1069–D1074 (2015).

40. Gottlieb, A., Stein, G. Y., Ruppin, E. & Sharan, R. PREDICT: a method for inferring novel drug indications with application to personalized medicine. *Mol. Syst. Biol.* **7**, 496 (2011).

41. Iwata, H., Sawada, R., Mizutani, S. & Yamanishi, Y. Systematic drug repositioning for a wide range of diseases with integrative analyses of phenotypic and molecular data. *J. Chem. Inf. Model.* **55**, 446–459 (2015).

42. Chen, H., Zhang, H., Zhang, Z., Cao, Y. & Tang, W. Network-based inference methods for drug repositioning. *Comput. Math. Methods M.* **2015**, 130620 (2015).

43. Luo, H. *et al.* Drug repositioning based on comprehensive similarity measures and Bi-Random walk algorithm. *Bioinformatics* **32**, 2664–2671 (2016).

44. Li, Y. *et al.* HMDDv2. 0: a database for experimentally supported human microRNA and disease associations. *Nucleic Acids Res.* **42**, D1070–D1074 (2013).

45. Chen, X. *et al.* NRDTD: a database for clinically or experimentally supported non-coding RNAs and drug targets associations. *Database* **2017**, bax057 (2017).

46. Wang, E. *et al.* Predictive genomics: a cancer hallmark network framework for predicting tumor clinical phenotypes using genome sequencing data. *Seminars in Cancer Biology* **30**, 4–12 (2015).

47. Hattori, M., Okuno, Y., Goto, S. & Kanehisa, M. Development of a chemical structure comparison method for integrated analysis of chemical and genomic information in the metabolic pathways. *J. Am. Chem. Soc.* **125**, 11853–11865 (2003).

48. Lv, S. *et al.* A novel method to quantify gene set functional association based on gene ontology. *J. R. Soc. Interface* **9**, 1063–1072 (2012).

49. Liu, X. *et al.* SM2miR: a database of the experimentally validated small molecules' effects on microRNA expression. *Bioinformatics* **29**, 409–411 (2012).

50. Chen, X., Liu, M.-X. & Yan, G.-Y. Drug–target interaction prediction by random walk on the heterogeneous network. *Mol. BioSyst.* **8**, 1970–1978 (2012).

51. Chen, X., Liu, M.-X. & Yan, G.-Y. RWRMDA: predicting novel human microRNA–disease associations. *Mol. BioSyst.* **8**, 2792–2798 (2012).

52. Chen, X., You, Z. H., Yan, G. Y. & Gong, D. W. IRWRLDA: improved random walk with restart for lncRNA-disease association prediction. *Oncotarget* **7**, 57919 (2016).

53. Chen, X. miREFRWR: a novel disease-related microRNA-environmental factor interactions prediction method. *Mol. BioSyst.* **12**, 624–633 (2016).

Acknowledgements

Author Contributions

H.C. and Z.Z. conceived and designed the experiments. H.C. performed the experiments. H.C., Z.Z. and W.P. analyzed the data. H.C. wrote the paper. All authors read and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-017-15716-8.

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.