



Article

A Novel Approach Based on a Weighted Interactive Network to Predict Associations of MiRNAs and Diseases

Haochen Zhao ², Linai Kuang ^{1,2}, Xiang Feng ^{1,2}, Quan Zou ^{3,4} and Lei Wang ^{1,2,*}

¹ College of Computer Engineering & Applied Mathematics, Changsha University, Changsha 410001, China

² Key Laboratory of Hunan Province for Internet of Things and Information Security, Xiangtan University, Xiangtan 411105, China; zhaohc940702@163.com (H.Z.); kuanglinai@xtu.edu.cn (L.K.); fengxiang@xtu.edu.cn (X.F.)

³ Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu 610000, China; zouquan@nclab.net

⁴ School of Computer Science and Technology, Tianjin University, Tianjin 300000, China

* Correspondence: wanglei@xtu.edu.cn; Tel.: +86-151-1110-9999

Received: 27 November 2018; Accepted: 24 December 2018; Published: 28 December 2018



Abstract: Accumulating evidence progressively indicated that microRNAs (miRNAs) play a significant role in the pathogenesis of diseases through many experimental studies; therefore, developing powerful computational models to identify potential human miRNA–disease associations is vital for an understanding of the disease etiology and pathogenesis. In this paper, a weighted interactive network was firstly constructed by combining known miRNA–disease associations, as well as the integrated similarity between diseases and the integrated similarity between miRNAs. Then, a new computational method implementing the newly weighted interactive network was developed for discovering potential miRNA–disease associations (WINMDA) by integrating the *T* most similar neighbors and the shortest path algorithm. Simulation results show that WINMDA can achieve reliable area under the receiver operating characteristics (ROC) curve (AUC) results of 0.9183 ± 0.0007 in 5-fold cross-validation, 0.9200 ± 0.0004 in 10-fold cross-validation, 0.9243 in global leave-one-out cross-validation (LOOCV), and 0.8856 in local LOOCV. Furthermore, case studies of colon neoplasms, gastric neoplasms, and prostate neoplasms based on the Human microRNA Disease Database (HMDD) database were implemented, for which 94% (colon neoplasms), 96% (gastric neoplasms), and 96% (prostate neoplasms) of the top 50 predicting miRNAs were confirmed by recent experimental reports, which also demonstrates that WINMDA can effectively uncover potential miRNA–disease associations.

Keywords: microRNA; diseases; association prediction; computational prediction model; path selection

1. Introduction

Recently, increasing studies indicated that non-coding RNAs (ncRNAs) play an extensive and important role in many biological processes such as cell differentiation, ontogenesis, and disease development [1–3]. In particular, microRNAs (miRNAs), a class of small ncRNAs with a length of 20–25 nucleotides, was proven to be closely related to the occurrence of many diseases that are seriously harmful to human health [4,5]; they are able to regulate many functions of eukaryotic cells and affect various behaviors such as gene expression, cell-cycle regulation, and individual development [6]. For example, miR-126 was demonstrated to be associated with clear cell human renal cell carcinoma [7], while miR-34a-5p was proven to have a critical impact on ovarian cancer (OC) cell

lines via interacting with nuclear paraspeckle assembly transcript 1 (NEAT1) [8]. MicroRNA expression microarray analysis showed that miR-145 and miR-1 expression is significantly downregulated in colon cancer tissues [9], and that miR-424 and miR-381 play important roles in tumor regulation, expression, and even treatment [10]. Therefore, it is necessary to study the association between miRNAs and diseases in depth and explore the potential relationship between miRNA and some human diseases [11,12].

The identification of potential miRNA–disease associations can not only play an important role in the diagnosis, treatment, and prevention of disease, but also effectively addresses the high cost and long-term shortcomings of traditional biological experiments [13,14]. Up to now, various miRNA-related heterogeneous biological databases were established and were extended to various fields of miRNA-related research, such as miRBase [15], Database of Differentially Expressed miRNAs in Human Cancers (dbDEMC) [16], Human microRNA Disease Database (HMDD) [17], miR2Disease [18], etc. Based on these datasets, different computational prediction methods were developed to predict potential miRNA–disease associations [19–24]. For example, in 2012, Chen et al. developed a prediction method named random walk with restart for miRNA–disease association (RWRMDA) for inferring potential miRNA–disease associations through combining the semantic similarity of the disease and the functional similarity of miRNA [25]. In 2014, Chen et al. further developed a computational model named regularized least squares for miRNA–disease association (RLSMDA) based on semi-supervised machine learning, which can uncover related miRNAs for each disease without any known related miRNAs [26]. In 2016, under the basic assumption that functional similar miRNAs tend to interact with similar diseases, Chen et al. proposed a prediction model named within and between Score for miRNA–disease association (WBSMDA) by combining the within score and between score from the opinion of diseases and miRNAs to calculate the score and predict potential miRNA–disease associations [27]. Furthermore, Zou et al. also developed a machine learning based method named CATAPULT to uncover relationships between microRNAs and diseases based on social network analysis methods [28]. Zhu et al. developed a novel path-based method named PBMMDA to predict the relationships between miRNA and disease by integrating different types of heterogeneous biological datasets and constructing three interlinked sub-graphs [29]. In 2017, Chen et al. proposed LRSSLMDA where novel miRNA–disease associations were predicted using sparse subspace learning to map high-dimensional miRNA/disease spaces into a lower-dimensional subspace [30]. In LRSSLMDA, feature extraction was performed on the integrated similarity to form the statistical profile and the graph theoretical profile. In 2018, Chen et al. developed a novel computational model named MDHGI to predict potential miRNA–disease associations using a sparse learning method to decompose the original adjacency matrix and combining the miRNA functional similarities network, the disease semantic similarities network, the Gaussian interaction profile kernel similarities network, and the new adjacency matrix into a heterogeneous graph [31]. Zhao and Wang presented a distance correlation set-based prediction method named DCSMDA for predicting latent miRNA–disease associations by constructing the disease/long non-coding RNA (lncRNA)/miRNA interactive network [32]. Moreover, Chen et al. developed an extreme gradient-boosting machine with a decision tree named EGBMMDA, seeking potential miRNA–disease associations using vector-covered statistical measures, graph theoretical measures, and matrix factorization results for miRNAs and diseases [33]. Chen et al. proposed a rating-integrated bipartite network recommendation-based prediction method named BNPMMDA for predicting potential miRNA–disease association using agglomerative hierarchical clustering [34].

In most of the abovementioned computational models, researchers constructed an adjacency matrix A to represent relationships between miRNA and disease. Specifically, if the association term between miRNA m_i and disease d_j is recorded in the database, the matrix $A(i,j) = 1$; otherwise it is equal to zero. However, since the number of known miRNA–disease associations existing in these well-known databases is very limited, this results in a sparse matrix; thus, in order to improve the accuracy of our prediction model, the information about the diseases and miRNAs

was adopted to construct a weighted interactive network in this paper. Moreover, on the basis of premises that functionally similar miRNAs may regulate similar diseases and similar diseases tend to associate with functionally similar miRNAs, a novel prediction model based on the newly constructed weighted interactive network was developed for miRNA–disease association inference (WINMDA). Comparing several state-of-the-art computational models, the strong point of WINMDA lies in the construction of a weighted interactive network and the introduction of the shortest path between nodes in the weighted interactive network. This also means that WINMDA proposes an idea for improving the sparseness of the adjacency matrix A and does not need negative samples to predict potential miRNA–disease associations simultaneously. All prediction results of potential miRNA–disease associations are shown in the Supplementary files 1–3; researchers could use these data to guide biological experiments in the future. In addition, the performance of WINMDA was evaluated by cross-validation and case studies of colon neoplasms, gastric neoplasms, and prostate neoplasms. Our simulation results show that WINMDA can achieve reliable area under the receiver operating characteristics (ROC) curve (AUC) results of 0.9183 ± 0.0007 , 0.9200 ± 0.0004 , 0.9243 , and 0.8856 in terms of 5-fold cross-validation, 10-fold cross-validation, global leave-one-out cross-validation (LOOCV), and local LOOCV, respectively. Additionally, 94% (colon neoplasms), 96% (gastric neoplasms), and 96% (prostate neoplasms) of the top 50 predicting miRNAs were confirmed by dbDEMC [16], miR2Disease [18], and recently published experimental studies. These results also demonstrate that WINMDA can effectively predict potential miRNA–disease associations.

2. Results and Case Studies

In this section, we evaluated the predictive performance of WINMDA through the following experiments: we firstly compared WINMDA with four state-of-the-art methods, namely BNPMDA [34], PBMDA [29], WBSMDA [27], and RLSMDA [26] in the framework of LOOCV. Then, the process of five-fold cross-validation was repeated 50 times for our method to evaluate the prediction performance of WINMDA. Thirdly, the influence of given parameters T and w on the prediction model was analyzed. Moreover, several case studies were performed to validate the feasibility of our method. Finally, experimental results regarding the top 50 predicted associations between miRNAs and four important neoplasms were listed, and we implemented the performance comparisons between WINMDA and four state-of-the-art methods through observing the rankings of six important disease-related miRNAs in the case studies.

2.1. Comparison with Existing State-of-the-Art Methods

We evaluated the performance of WINMDA by observing it along with four state-of-the-art methods to predict the accuracy of potential miRNA–disease associations using global and local LOOCV. In the global LOOCV, each known disease–miRNA association was alternately used as a test sample, and other known miRNA–disease associations were considered as a training set, while all unknown disease–miRNA associations in HMDD were considered as a candidate set. However, in local LOOCV, for each given disease d , each known miRNA related to d was utilized as a test sample, while all other known miRNAs related to d were used as training samples, and all other unknown miRNAs related to d were considered as candidates. Hence, through comparing the scores obtained from the test sample with other candidate associations, we could evaluate how well this association was ranked in the candidate set; if the predicted ranking of the test sample was higher than the preset threshold, then the sample was successfully predicted by the computational model. In other words, let TP be true positive, TN be true negative, FN be false negative, and FP be false positive; then, under different threshold settings, the corresponding true positive rate (TPR; sensitivity) and false positive rate (FPR; specificity) can be obtained as follows:

$$\text{FPR} = \text{FP}/(\text{FP} + \text{TN}) \quad (1)$$

$$\text{TPR} = \text{TP}/(\text{TP} + \text{FN}) \quad (2)$$

Here, sensitivity means that the percentage of the test samples with predicted ranks was higher than the given threshold, whereas specificity was computed as the percentage of negative samples with predicted ranks lower than the given threshold. Obviously, after obtaining different TPR and FPR pairs under different thresholds, the final ROC curve (in which, FPR is the horizontal axis of the coordinate system and TPR is the longitudinal axis of the coordinate system) could be plotted by connecting these pairs. Finally, the AUC could be obtained to represent the specific prediction performance of the computational models. Furthermore, it is obvious that the larger the AUC value is, the more likely the current classification algorithm is to place a positive sample in front of the negative sample, such that it can be better classified. Next, for the global LOOCV and local LOOCV, we compared WINMDA with four state-of-the-art computational methods, namely BNPMDA, PBMDA, WBSMDA, and RLSMDA; the simulation results are shown in Figures 1 and 2, respectively. From the two figures, it is easy to see that WINMDA can achieve reliable AUCs of 0.9243 and 0.8856 for global LOOCV and local LOOCV, respectively, when $T = 16$ and $w = 0.6$, which are much higher than the AUCs of 0.9169 and 0.8523 achieved by PBMDA, 0.9082 and 0.8571 achieved by BNPMDA, 0.8030 and 0.8390 achieved by WBSMDA, and 0.8426 and 0.7169 achieved by RLSMDA. It is obvious that our newly proposed method WINMDA is superior to these four traditional computational models in global and local LOOCV; therefore, it can be used as an important tool for discovering potential miRNA–disease associations.

Table 1. The average area under the receiver operating characteristics (ROC) curve (AUC) achieved by the weighted interactive network for miRNA–disease association inference (WINMDA) under the frameworks of 5-Fold cross-validation and 10-Fold cross-validation.

LOOCV	5-Fold Cross-Validation	10-Fold Cross-Validation
0.9243	0.9183 ± 0.0007	0.9200 ± 0.0004

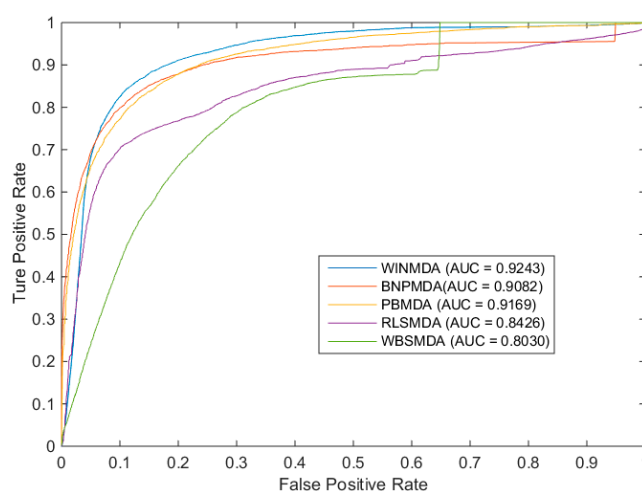


Figure 1. The comparison results between the weighted interactive network for miRNA–disease association inference (WINMDA) and four state-of-the-art computational models in terms of global leave-one-out cross-validation (LOOCV).

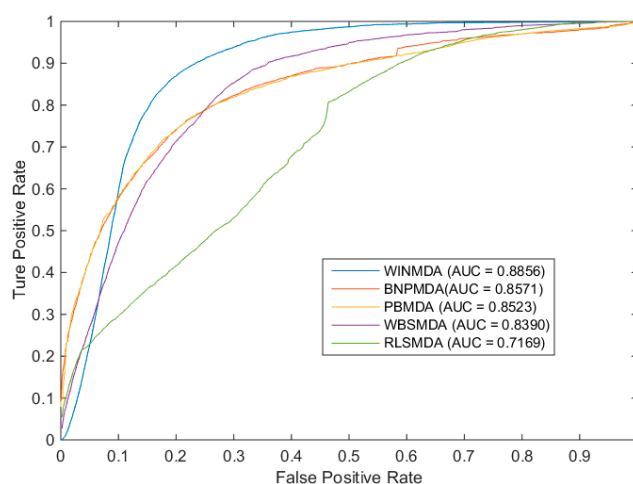


Figure 2. The comparison results between WINMDA and four state-of-the-art computational models in terms of local LOOCV.

In addition, considering the potential bias of random sample partitioning for performance assessment, we divide the known miRNA-disease associations by 50 times, and the AUCs were obtained in the similar way as global LOOCV. As a result, WINMDA achieved the prediction performance with average AUCs of 0.9183 and 0.9200 with standard deviation of 0.0007 and 0.0004 when using the 5-Fold and 10-Fold cross validation (Table 1).

2.2. Evaluation of the Effects of Parameters

There are two kinds of important parameters existing in our newly proposed model of WINMDA, as illustrated in Equation (22), one is w and the other is T .

2.2.1. Effects of Parameter T

From Equations (20), (21), and (22), it is easy to know that parameter T will have important effects on the accuracy of our prediction model WINMDA. For instance, if the value of T is too large, then lots of noise data will be included, which will reduce the predictive performance of WINMDA. Alternatively, if the value of T is too small, then the useful associations may not be sufficient for accurate estimation of potential associations between some specific diseases and miRNAs. Hence, in order to evaluate the effects of parameter T , we set the value of T ranging from 1–20 during the implementation of WINMDA, and the simulation results are shown in Table 2. From Table 2, it is easy to see that the AUCs achieved by WINMDA varied with the different values of T . Specifically, the prediction performance of WINMDA increased upon increasing of the value of T , while T varied from 1 to 16, which indicates that the number of useful neighbors is positively related to the prediction performance of our model. Meanwhile, it is easy to find that the AUC will decline when $T > 16$, which indicates that an excess of noise data will markedly interfere with our prediction model WINMDA. Therefore, it was determined as best to set T to 16 for WINMDA in this paper.

Table 2. Effects of T on the prediction performance of WINMDA when $w = 0.6$.

T	AUC	T	AUC
1	0.9145	12	0.9241
2	0.9160	16	0.9243
5	0.9222	18	0.9242
8	0.9244	20	0.9188

2.2.2. Effects of Parameter w

In this section, to investigate the effects of parameter w on the prediction performance of WINMDA, we set w to different values ranging from 0–1, while implementing WINMDA under LOOCV, and the results are shown in the Table 3. It is obvious that the variation of the value of w has an important influence on the performance of our prediction model WINMDA. Specifically, from Table 3, it is obvious that WINMDA can achieve the maximum AUC value when w is set to 0.6. Hence, we set w to 0.6 in this paper.

Table 3. Effects of w on the prediction performance of WINMDA when $T = 16$.

w	AUC	w	AUC
0	0.9135	0.6	0.9243
0.1	0.9188	0.7	0.9239
0.2	0.9209	0.8	0.9222
0.3	0.9216	0.9	0.9189
0.4	0.9235	1	0.9160
0.5	0.9241		

2.3. Case Studies

Recently, increasing evidence demonstrated that miRNAs play an extensive and important role in the physiological processes of the body [35]. In addition, in developed countries such as the United States and throughout Europe, cancer is the second leading cause of human death, while it ranks second or third in developing countries [36]. Therefore, in order to further evaluate the accuracy of WINMDA in predicting potential disease–miRNA associations, we chose three kinds of cancers, i.e., colon neoplasms, gastric neoplasms, and prostate neoplasms, as case studies for WINMDA, and the prediction results were verified by recently published experimental studies and two databases, namely miR2Disease and dbDEMC. During the simulation, for each kind of cancer, all known related miRNAs were considered as seed miRNAs, and the other miRNAs were considered as candidate miRNAs. In addition, all candidate miRNAs associated with colon neoplasms, gastric neoplasms, and prostate neoplasms were ranked in descending order according to our prediction results, as illustrated in Tables 4–6, respectively.

Recently, colon cancer (colon neoplasms) ranks third among the most common female cancers and second among the most common male cancers in the world [37]. Each year, more than one million people died from colon cancer [38]. The incidence rates vary widely around the world, depending on lifestyle, environment, and heredity [39]. Recent studies reported that miRNAs are closely related to the diagnosis, prognosis, and chemo-sensitivity of colon cancer, which indicates that miRNAs can be used as a marker for the early diagnosis of colon cancer and as a guideline for various stages of colon cancer [40]. Hence, case studies on colon cancer-related miRNAs were implemented to further verify the predictive ability of WINMDA and, as a result, 10 of the top 10 and 47 of the top 50 candidate miRNAs were shown to be associated with colon neoplasms by miR2Disease, dbDEMC, and other known experimental studies (Table 4). For example, some researchers confirmed that the overexpression of miR-143 (ranked first in the WINMDA forecast list) reduces cell proliferation and migration of mutant *KRAS* HCT116 colon cancer cells [41]. Additionally, experimental studies also found that miR-20a is a member of the miR-17 miRNA family, which is part of the regulatory machinery that defines the pro-tumorigenic differentiation of stromal fibroblasts. In stromal fibroblasts, miR-20a (ranked second in the WINMDA forecast list) can modulate chemokine C–X–C ligand 8 (CXCL8) function, thereby influencing tumor latency [42]. Moreover, some researchers found that the rs35301225 polymorphism in miR-34a (ranked third in the WINMDA forecast list) is involved in the development of human colon cancer via downregulation of tumor-promoting gene *E2F1* as a tumor suppressor, and the C/A single-nucleotide polymorphism of miR-34a promotes colon cancer cell proliferation via upregulating *E2F1* [43].

In recent years, it was reported that gastric cancer (gastric neoplasms) is one of the most common malignant tumors of the digestive tract in the world, and Japan, South Korea, and China are high-risk areas for gastric cancer [44]. Therefore, it is necessary to explore the mechanism of miRNA in the development of gastric cancer and provide a basis for the early diagnosis of gastric cancer. We used potential gastric cancer-associated miRNAs as a case study to further illustrate the predictive power of WINMDA in this section. As a result, 10 of the top 10 and 48 of the top 50 potential gastric cancer-related miRNAs were validated by miR2Disease, dbDEMC, and other known experimental studies (Table 5). For example, gene *UHRF1* plays a significant role in the development of gastric cancer. Furthermore, Zhou et al. identified and verified miR-146b (ranked first in our prediction list) and miR-146a as direct upstream regulators of *UHRF1* in gastric cancer metastasis. [45]. In addition, according to the target genes of miR-143 (ranked seventh in our prediction list), *IGF1R* and *BCL2*, which are related to cisplatin resistance, we can regulate the resistance of human gastric cancer cells to cisplatin via differential expression of *IGF1R* and *BCL2* in gastric cancer tissues and cell lines [46].

Prostate cancer (prostate neoplasms) is the third most common type of cancer. In 2012, the incidence rate of prostate cancer in the neoplasm registration area in China was 99.2%, which ranked sixth in the incidence of male malignant tumors [47]. However, early patients with prostate tumors have only subtle symptoms that make it difficult to detect cancer at an early stage [48]. Increasing studies confirmed that some miRNAs are related to prostate neoplasms. Therefore, case studies about prostate cancer-related miRNAs were implemented to further verify the predictive ability of WINMDA in this section. As a result, nine of the top 10 and 48 of the top 50 predicted prostate cancer-related miRNAs were validated by miR2Disease, dbDEMC, and other known experimental studies (Table 6). For example, Chu et al. selected single-nucleotide polymorphisms (SNPs) in the 1000 bp upstream from the transcription start site of hsa-miR-143 (ranked first in our prediction list) precursor in the dbSNP database with the condition that MAF > 0.05 in the Chinese population and finally identified that rs4705342 T > C was associated with the risk of prostate cancer, and that the C allele had a protective effect [49]. Wang et al. explored the effects of miR-182 (ranked second in our prediction list) on the growth, migration, and apoptosis of prostate cancer cells using qRT-PCR analysis. Moreover, they found that miR-182 plays an important role in prostate cancer, which enhances HIF1 α signaling by targeting PHD2 and FIH1 in prostate cancer [50]. Furthermore, Taddei et al. confirmed that hsa-miR-210 (ranked fourth in our prediction list) overexpression increased senescence-associated features in young fibroblasts and converted them into cancer-associated fibroblast-like cells. These senescent fibroblasts can induce epithelial–mesenchymal transition in prostate cancer cells, support tumor angiogenesis, and recruit endothelial precursor cells, thus contributing to cancer progression [51].

In order to further illustrate the high efficiency of our method, we compared the performances of PBMDA, WBSMDA, LRLSMDA, and our model WINMDA by counting the top 50 disease-associated miRNAs in the predicted results and observing how many disease-related miRNAs were identified by miR2Disease, dbDEMC, and recent biological experimental studies (Table 7) in the case studies of six important diseases. As a result, from Table 7, it is easy to see that WINMDA is more effective than other methods in general. In addition, as a global computational model, WINMDA can not only achieve reliable prediction performances, but also simultaneously predict all potential associations between the diseases and miRNAs in HMDD, which means that potential associations with high predicted values obtained by WINMDA can be used preferentially for biological experiment verification and public release. Hence, we may easily reach a conclusion that our newly proposed model WINMDA is of great value in predicting potential miRNA–disease associations.

Table 4. The potential top 50 predicted microRNAs (miRNAs) related to colon neoplasms obtained by WINMDA based on known associations in the Human microRNA Disease Database (HMDD) database.

Top 1–25 miRNAs	Evidence	Top 26–50 miRNAs	Evidence
hsa-mir-143	dbDEMC and miR2Disease	hsa-let-7e	dbDEMC
hsa-mir-20a	dbDEMC and miR2Disease	hsa-mir-486	26895105
hsa-mir-34a	dbDEMC and miR2Disease	hsa-mir-133b	dbDEMC and miR2Disease
hsa-mir-210	dbDEMC	hsa-mir-200a	unconfirmed
hsa-mir-21	dbDEMC and miR2Disease	hsa-mir-141	dbDEMC and miR2Disease
hsa-mir-155	dbDEMC and miR2Disease	hsa-let-7f	dbDEMC and miR2Disease
hsa-mir-95	dbDEMC and miR2Disease	hsa-mir-29a	dbDEMC and miR2Disease
hsa-mir-146a	dbDEMC	hsa-mir-181a	dbDEMC and miR2Disease
hsa-mir-16	dbDEMC	hsa-mir-9	dbDEMC and miR2Disease
hsa-mir-125b	dbDEMC	hsa-mir-29b	dbDEMC and miR2Disease
hsa-mir-92a	unconfirmed	hsa-let-7c	dbDEMC
hsa-mir-31	dbDEMC and miR2Disease	hsa-let-7d	dbDEMC
hsa-mir-223	dbDEMC and miR2Disease	hsa-mir-196a	dbDEMC and miR2Disease
hsa-mir-221	dbDEMC and miR2Disease	hsa-let-7i	dbDEMC
hsa-mir-222	dbDEMC	hsa-mir-142	23619912
hsa-let-7a	dbDEMC and miR2Disease	hsa-mir-1	dbDEMC and miR2Disease
hsa-mir-19b	dbDEMC and miR2Disease	hsa-mir-133a	dbDEMC and miR2Disease
hsa-mir-15a	dbDEMC	hsa-mir-192	dbDEMC and miR2Disease
hsa-mir-18a	dbDEMC and miR2Disease	hsa-mir-150	26455323
hsa-mir-200b	dbDEMC	hsa-mir-203	dbDEMC and miR2Disease
hsa-mir-19a	dbDEMC and miR2Disease	hsa-mir-451a	25484364
hsa-let-7b	dbDEMC and miR2Disease	hsa-let-7g	dbDEMC and miR2Disease
hsa-mir-24	miR2Disease	hsa-mir-124	dbDEMC
hsa-mir-199a	unconfirmed	hsa-mir-224	dbDEMC and miR2Disease
hsa-mir-200c	dbDEMC and miR2Disease	hsa-mir-146b	28466779

Table 5. The potential top 50 predicted miRNAs related to gastric neoplasms obtained by WINMDA based on known associations in the HMDD database.

Top 1–25 miRNAs	Evidence	Top 26–50 miRNAs	Evidence
hsa-mir-146b	26673617	hsa-mir-20a	29450946
hsa-mir-130a	25834316	hsa-mir-375	21343377
hsa-mir-21	miR2Disease	hsa-mir-17	30024601
hsa-mir-146a	28922434	hsa-mir-222	miR2Disease
hsa-mir-155	26950485	hsa-mir-101	28944848
hsa-mir-145	miR2Disease	hsa-mir-199a	24655788
hsa-mir-143	miR2Disease	hsa-mir-22	28482669
hsa-mir-200a	25740983	hsa-mir-196a	24527072
hsa-mir-200b	25740983	hsa-mir-223	22270966
hsa-mir-126	26464628	hsa-mir-7	26261179
hsa-mir-200c	27766962	hsa-mir-34c	18803879
hsa-let-7a	miR2Disease	hsa-mir-122	29509059
hsa-mir-141	miR2Disease	hsa-mir-218	27696291
hsa-mir-34a	25834316	hsa-mir-34b	unconfirmed
hsa-mir-142	21343377	hsa-mir-10b	25190020
hsa-mir-31	19598010	hsa-mir-103a	29754469
hsa-mir-16	miR2Disease	hsa-mir-27a	miR2Disease
hsa-mir-192	24981590	hsa-mir-150	20067763
hsa-mir-486	26895105	hsa-mir-18a	26950485
hsa-mir-221	miR2Disease	hsa-mir-19a	22802949
hsa-mir-107	miR2Disease	hsa-mir-106a	miR2Disease
hsa-let-7f	21533124	hsa-mir-9	28418879
hsa-let-7g	25972194	hsa-mir-451a	unconfirmed
hsa-mir-133b	23296701	hsa-mir-124	27041578
hsa-mir-125b	24846940	hsa-mir-1	25874496

Table 6. The potential top 50 predicted miRNAs related to prostate neoplasms obtained by WINMDA based on known associations in the HMDD database.

Top 1–25 miRNAs	Evidence	Top 26–50 miRNAs	Evidence
hsa-mir-143	dbDEMCM and miR2Disease	hsa-mir-15a	dbDEMCM and miR2Disease
hsa-mir-182	dbDEMCM and miR2Disease	hsa-mir-181b	dbDEMCM and miR2Disease
hsa-mir-96	dbDEMCM and miR2Disease	hsa-mir-375	dbDEMCM and miR2Disease
hsa-mir-34a	dbDEMCM and miR2Disease	hsa-mir-200a	dbDEMCM
hsa-mir-210	miR2Disease	hsa-mir-34b	dbDEMCM
hsa-mir-150	dbDEMCM	hsa-mir-34c	dbDEMCM
hsa-mir-92a	Unconfirmed	hsa-let-7b	dbDEMCM and miR2Disease
hsa-mir-141	miR2Disease	hsa-mir-218	dbDEMCM and miR2Disease
hsa-mir-21	dbDEMCM and miR2Disease	hsa-mir-101	dbDEMCM and miR2Disease
hsa-mir-222	dbDEMCM and miR2Disease	hsa-mir-124	dbDEMCM
hsa-mir-31	dbDEMCM and miR2Disease	hsa-mir-223	dbDEMCM and miR2Disease
hsa-mir-146b	25712341	hsa-let-7a	dbDEMCM and miR2Disease
hsa-mir-221	dbDEMCM and miR2Disease	hsa-mir-224	dbDEMCM and miR2Disease
hsa-mir-203	26499781	hsa-mir-205	dbDEMCM and miR2Disease
hsa-mir-126	dbDEMCM and miR2Disease	hsa-let-7d	dbDEMCM and miR2Disease
hsa-mir-200b	Unconfirmed	hsa-mir-1	dbDEMCM
hsa-mir-200c	dbDEMCM	hsa-let-7c	dbDEMCM and miR2Disease
hsa-mir-146a	miR2Disease	hsa-mir-127	dbDEMCM and miR2Disease
hsa-mir-17	miR2Disease	hsa-mir-135b	dbDEMCM
hsa-mir-100	dbDEMCM and miR2Disease	hsa-mir-214	dbDEMCM and miR2Disease
hsa-mir-16	dbDEMCM and miR2Disease	hsa-mir-93	26124181
hsa-mir-199a	dbDEMCM and miR2Disease	hsa-mir-708	22552290
hsa-mir-20a	miR2Disease	hsa-mir-155	dbDEMCM
hsa-mir-133b	dbDEMCM	hsa-mir-133a	dbDEMCM
hsa-mir-27b	dbDEMCM and miR2Disease	hsa-mir-195	dbDEMCM and miR2Disease

Table 7. Effects of w on the prediction performance of WINMDA when $T = 16$.

Disease	WINMDA	BNPMDA	PBMDA	WBSMDA	RLSMDA
Breast neoplasms	44	48	46	36	42
Colon neoplasms	47	45	47	45	46
Gastric neoplasms	48	43	46	43	44
Kidney neoplasms	45	43	42	42	45
Liver neoplasms	48	45	45	46	46
Prostate neoplasms	48	44	45	42	44

3. Discussion

Increasing studies based on biological experiments indicated that miRNAs are closely related to the occurrence of many diseases that are seriously harmful to human health, and the identification of potential miRNA–disease associations can not only play an important role in the diagnosis, treatment, and prevention of disease, but also effectively addresses the high cost and long-term shortcomings of traditional biological experiments. In this article, we developed a novel prediction model called WINMDA to predict potential relationships between miRNAs and diseases based on premises that functionally similar miRNAs may regulate similar diseases and similar diseases tend to associate with functionally similar miRNAs. In WINMDA, we firstly integrated disease semantic similarity, Gaussian interaction profile kernel similarity, and miRNA function similarity, and then constructed a weighted interactive network for potential miRNA–disease prediction. The important difference between WINMDA and previous state-of-the-art prediction models is that the problem of limited known miRNA–disease associations was considered in WINMDA and the shortest paths in the weighted interactive network were adopted to solve the problem. Moreover, we evaluated the predictive performance of WINMDA through LOOCV (including global LOOCV and local LOOCV), k -fold cross-validation, and several case studies. Experimental results show that WINMDA can

effectively uncover potential disease–miRNA candidates, which means that it can be used as a reliable and accurate calculation model for finding potential miRNA–disease associations.

Although WINMDA achieved effective performance in predicting candidate relationships between miRNAs and diseases, there are still some existing limitations that can be improved in the future. Firstly, the parameters T and w play important roles in WINMDA, and the selection of suitable values for T and w are critical problems that shall be addressed in future studies. Secondly, the assigned weight may not be accurate enough, as it was on the basis of premises that functionally similar miRNAs may regulate similar diseases and similar diseases tend to associate with functionally similar miRNAs. Finally, a weighted interactive network was constructed in WINMDA based on the disease similarity, miRNA similarity, fig:ijms-405460-f003 and known miRNA–disease associations. The performance of WINMDA will be further improved considering more databases storing other information about diseases, miRNAs, and miRNA–disease associations.

4. Materials and Methods

4.1. Construction of the miRNA–Disease Interactive Network

In order to construct the miRNA–disease interactive network, we firstly downloaded known miRNA–disease associations from the HMDD database on 14 July 2018. After eliminating duplicate values, erroneous data, and disorganized data, human miRNA–disease associations were downloaded from the HMDD database, which includes 5430 experimentally validated human miRNA–disease associations involving 495 miRNAs and 383 diseases. (Supplementary file 4) have been collected. Let D represent the number of different disease items and M represent the number of different miRNA items in the HMDD database, respectively; let $S_D = \{d_1, d_2, \dots, d_D\}$ represent the set of these D different diseases, and $S_M = \{m_{D+1}, m_{D+2}, \dots, m_{D+M}\}$ represent the set of these M different miRNAs. Then, we can construct an miRNA–disease interactive network $G = (V, E)$, where $V = S_D \cup S_M = \{d_1, d_2, \dots, d_D, m_{D+1}, m_{D+2}, \dots, m_{D+M}\}$ is the set of vertices, E is the edge set of G , and $\forall d_i \in S_D, m_j \in S_M$. There is an edge between d_i and m_j in E if and only if there is an association between m_j and d_i in the database of HMDD. Thereafter, based on the newly constructed miRNA–disease interactive network G , for any given $d_i \in S_D$ and $m_j \in S_M$, we can obtain a $D \times M$ dimensional matrix DMM as follows:

$$DMM(i, j) = \begin{cases} 1 & \text{if } d_i \text{ is related to } m_j \text{ in HMDD} \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

4.2. Calculation of the Disease Semantic Similarity

For any two diseases d_i and d_j that belong to S_D , the semantic similarity between d_i and d_j was calculated according to the following steps:

Step 1: Firstly, we collected the Medical Subject Headings (MeSH) descriptors of d_i and d_j from the National Library of Medicine (<http://www.nlm.nih.gov/>).

Step 2: Secondly, we constructed direct acyclic graphs (DAGs) corresponding to d_i and d_j separately and, as illustrated in the Figure 3, for any given disease H , its DAG can be represented as $DAG(H) = (N(H), E(H))$, where $N(H)$ is the node set and $E(H)$ is the edge set. Moreover, in $DAG(H)$, each node corresponds to a different disease MeSH descriptor, and all the MeSH descriptors are connected by a direct edge from a more general term (called a parent node) to a more specific term (called a child node). Furthermore, $N(H)$ consists of the node H itself and its ancestor nodes; $E(H)$ consists of corresponding direct edges from a parent node to a child node, and each edge in $E(H)$ represents the relationship between these two nodes connected by it.

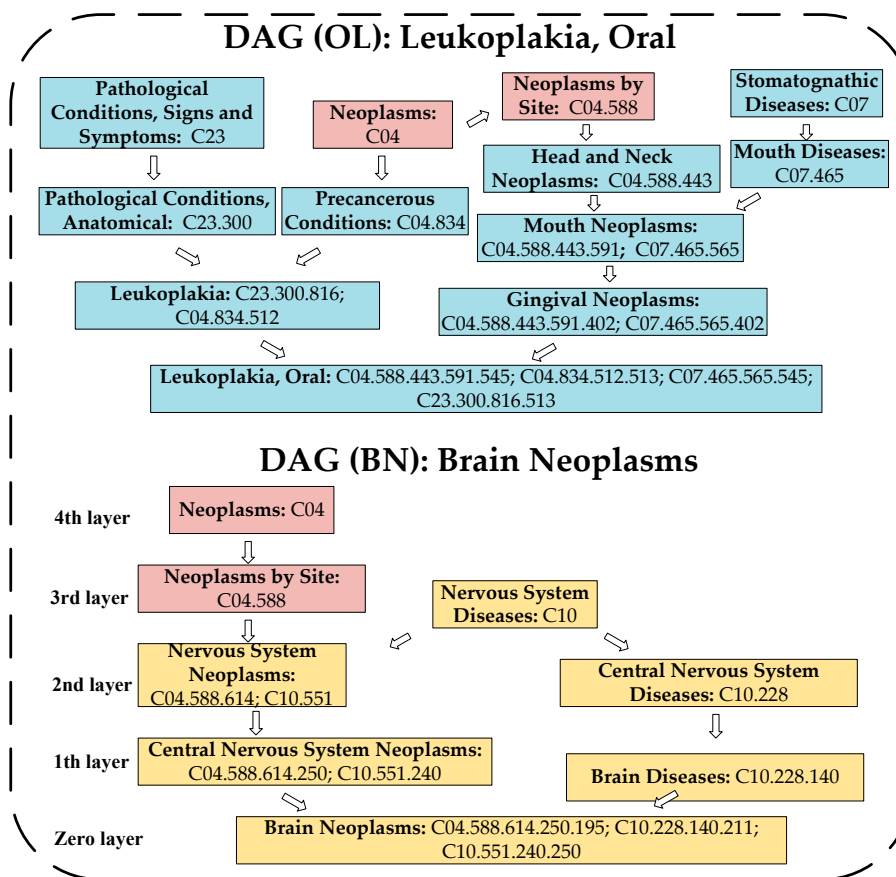


Figure 3. The disease directed acyclic graphs (DAGs) of leukoplakia, and oral and brain neoplasms.

Step 3: Thirdly, based on the newly constructed $DAG(H)$, let d be an ancestor node of H in $DAG(H)$; then, we defined the contribution of an ancestor node d to the semantic value of the disease H and the contribution of the semantic value of disease H itself as follows:

$$\begin{cases} D_H(d) = 1 & \text{if } d = H; \\ D_H(d) = \sum\{\alpha \times \beta \times D_H(d^*) | d^* \in \text{children of } d\} & \text{if } d \neq H. \end{cases} \quad (4)$$

Here, the parameter α is a semantic contribution attenuation factor with a value between zero and one, and its value was set to 0.5 in this paper according to previous state-of-the-art methods [52,53]. The parameter β is the number of addresses or codes included in the node d^* , which indicates the weight of the contribution of disease d^* for H in $DAG(H)$.

Obviously, according to Equation (2), it is easy to know that an ancestor node d with a larger number of child nodes in $DAG(H)$ will make a more significant contribution to the semantic value of H . For instance, in $DAG(BN)$ of Figure 3, the entry on “central nervous system neoplasms” includes two addresses or codes: C04.588.614.250 and C10.551.240; however, the entry on “brain diseases” includes only one code: C10.228.140. Thus, the contribution of “central nervous system neoplasms” to the semantic value of the “brain neoplasms” is $2 \times \alpha \times 1$, while the contribution of “brain diseases” to the semantic value of the “brain neoplasms” is $1 \times \alpha \times 1$ only.

Step 4: Next, based on Equation (2), we calculated the semantic value of disease H by accumulating the contributions of all disease terms to H in $DAG(H)$ as follows:

$$TS(H) = \sum_{d \in S(H)} D_H(d) \quad (5)$$

For example, according to Equation (3), in DAG(BN) of Figure 3, the semantic value of the disease “brain neoplasms” can be obtained by “the contributions of ‘brain neoplasms’ to it” ($= 1 \times 1$) + “the contributions of ‘central nervous system neoplasms’ to it” ($= 2 \times 0.5$) + “the contributions of ‘brain diseases’ to it” ($= 1 \times 0.5$) + “the contributions of ‘nervous system neoplasms’ to it” ($= 2 \times 0.5 \times 0.5$) + “the contributions of ‘central nervous system diseases’ to it” ($= 1 \times 0.5 \times 0.5$) + “the contributions of ‘nervous system diseases’ to it” ($= 1 \times 0.5 \times 0.5 \times 0.5 + 1 \times 0.5 \times 0.5 \times 0.5$) + “the contributions of ‘neoplasms by site’ to it” ($= 1 \times 0.5 \times 0.5 \times 0.5$) + “the contributions of ‘neoplasms’ to it” ($= 1 \times 0.5 \times 0.5 \times 0.5 \times 0.5$) = 3.6875.

Step 5: Finally, we defined the semantic similarity between d_i and d_j as follows:

$$SDD(d_i, d_j) = \frac{\sum_{d \in S(d_i) \cap S(d_j)} (D_{d_i}(d) + D_{d_j}(d))}{TS(d_i) + TS(d_j)} \tag{6}$$

Additionally, for any two diseases d_a and d_b that belong to S_D , if d_a and d_b do not have semantic similarity, we define $SDD(d_a, d_b) = -1$; then, based on Equation (4), it is obvious that we can obtain a $D \times D$ dimensional disease semantic similarity matrix $SDD(i, j)$.

4.3. Calculation of the miRNA Functional Similarity

Considering that, in the HMDD database, one miRNA may be associated with multiple disease items and vice versa, and, according to the state-of-the-art literature [39], the functional similarity can be obtained by integrating the semantic similarity of the two groups of diseases associated with these two miRNAs, then, for any two diseases m_i and m_j that belong to S_M , the functional similarity between m_i and m_j can be calculated according to the following steps:

Step 1: Firstly, let dx be any given disease, and $Dgroup = \{dy_1, dy_2, dy_3 \dots dy_r\}$ be a set consisting of r different diseases, and then the semantic similarity between dx and $Dgroup$ can be calculated as follows:

$$SS(dx, Dgroup) = \max_{1 \leq i \leq r} (SDD(dx, dy_i)) \tag{7}$$

For example, let d_a, d_b , and d_c be three kinds of diseases, $Dgroup = \{d_b, d_c\}$, $SDD(d_a, d_b) = 0.7$, and $SDD(d_a, d_c) = 0.8$; then, the semantic similarity between d_a and $Dgroup$ is $SS(d_a, Dgroup) = \max\{SDD(d_a, d_b), SDD(d_a, d_c)\} = 0.8$.

Step 2: Secondly, let $Dgroup_i$ and $Dgroup_j$ be the sets of diseases associated with m_i and m_j , respectively; supposing that there are N and M different diseases in $Dgroup_i$ and $Dgroup_j$, then we calculated the functional similarity between m_i and m_j as follows:

$$SMM(m_i, m_j) = \frac{\sum_{1 \leq i \leq M} SS(d_i, Dgroup_i) + \sum_{1 \leq j \leq N} SS(d_j, Dgroup_j)}{M + N}, \tag{8}$$

where $d_i \in Dgroup_j$ and $d_j \in Dgroup_i$. For example, let $Dgroup_1 = \{X, Y\}$, $Dgroup_2 = \{X, Z\}$, supposing that $SDD(X, Y) = 0.6$, $SDD(X, Z) = 0.7$, and $SDD(Y, Z) = 0.5$, then $MM(m_1, m_2)$ can be obtained as follows:

$$\begin{aligned} SMM(m_1, m_2) &= \frac{SS(X, Dgroup_1) + SS(Y, Dgroup_2) + SS(X, Dgroup_1) + SS(Z, Dgroup_2)}{2+2} \\ &= \frac{SDD(X, X) + SDD(Y, X) + SDD(X, X) + SDD(Z, X)}{4} = \frac{1+0.6+1+0.7}{4} = 0.825. \end{aligned}$$

Additionally, for any two miRNAs m_a and m_b that belong to S_M , if m_a and m_b do not have semantic similarity, we define $SMM(m_a, m_b) = -1$; then, based on Equation (6), it is obvious that we can obtain an $M \times M$ dimensional miRNA functional similarity matrix $SMM(i, j)$.

4.4. Disease Gaussian Interaction Profile Kernel Similarity Measurement

On the basis of premises that functionally similar miRNAs may regulate similar diseases and similar diseases tend to associate with functionally similar miRNAs, let $DLP(d_i)$ represent the i -th row in the matrix DMM ; then, for any two diseases d_i and d_j that belong to S_D , we can calculate the Gaussian interaction profile kernel similarity between them as follows:

$$DGS(d_i, d_j) = \exp\left(-\frac{D \times \|DLP(d_i) - DLP(d_j)\|^2}{\sum_{i=1}^D \|DLP(d_i)\|^2}\right) \quad (9)$$

Additionally, based on previous work [54], we can further improve the disease Gaussian interaction profile kernel similarity using a logistic function as follows:

$$FDGS(i, j) = \frac{1}{1 + e^{-15 * DGS(d_i, d_j) + \log(9999)}} \quad (10)$$

4.5. MicroRNA Gaussian Interaction Profile Kernel Similarity Measurement

On the basis of premises that functionally similar miRNAs may regulate similar diseases and similar diseases tend to associate with functionally similar miRNAs, let $MLP(m_i)$ represent the i -th column in the matrix DMM ; then, for any two miRNAs m_i and m_j that belong to S_M , we can calculate the Gaussian interaction profile kernel similarity between them as follows:

$$FMGS(i, j) = \exp\left(-\frac{M * \|MLP(m_i) - MLP(m_j)\|^2}{\sum_{i=1}^M \|MLP(m_i)\|^2}\right). \quad (11)$$

4.6. Calculation of the Integrated Similarity

Based on Equations (6) and (10), the disease integrated similarity matrix FDD can be calculated based on the disease semantic similarity matrix (SDD) and the disease Gaussian interaction profile kernel similarity matrix ($FDGS$) as follows:

$$FDD(i, j) = \begin{cases} SDD(i, j) & \text{if } SDD(i, j) \geq 0, \\ FDGS(i, j) & \text{otherwise.} \end{cases} \quad (12)$$

Similarly, based on Equations (8) and (11), the miRNA integrated similarity matrix FMM can be calculated based on the miRNA functional similarity matrix (SMM) and the miRNA Gaussian interaction profile kernel similarity matrix ($FMGS$) as follows:

$$FMM(i, j) = \begin{cases} SMM(i, j) & \text{if } SMM(i, j) \geq 0, \\ FMGS(i, j) & \text{otherwise.} \end{cases} \quad (13)$$

4.7. Construction of the Weighted Interactive Network

For any given miRNA $m_i \in S_M$, we define the miRNA $m_x \in S_M$ as the most related miRNA to m_i , if m_x satisfies the following:

$$FMM(m_x, m_i) = \max_{1 \leq l \leq M} (FMM(m_l, m_i)), \text{ where } m_x \neq m_i. \quad (14)$$

Thereafter, as illustrated in Figure 4, we can construct the weighted interactive network according to the following four steps:

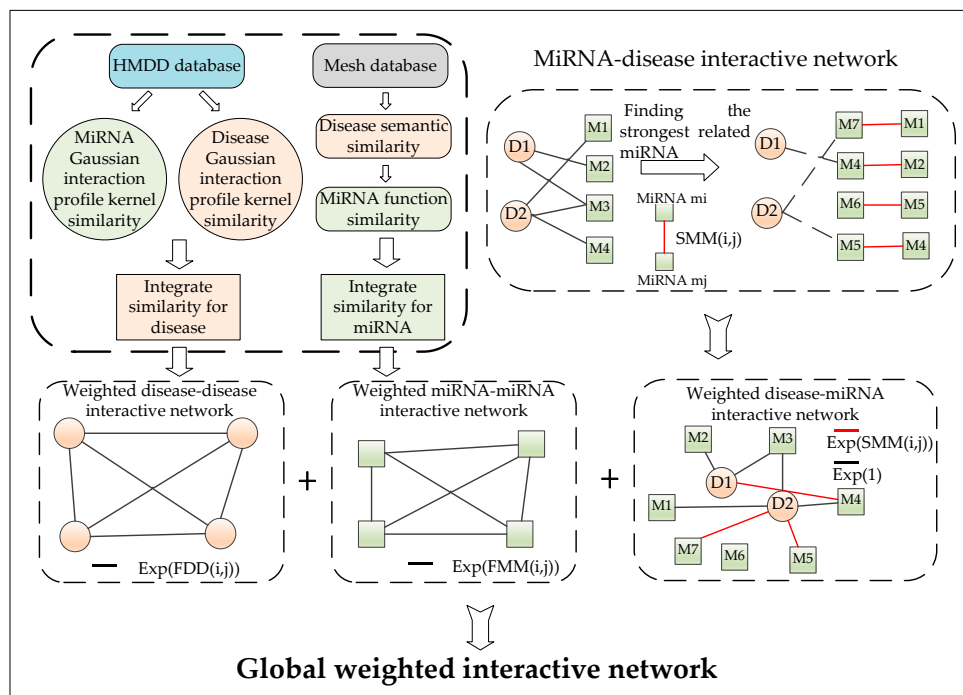


Figure 4. The flowchart detailing the construction of the global weighted interactive network by combining the weighted disease–disease interactive network, the weighted miRNA–miRNA interactive network, and the weighted disease–miRNA interactive network.

Step1: Firstly, for any given disease $d_i \in S_D$, we define the miRNA m_j as a potential miRNA to d_i if and only if m_j satisfies $DMM(i,j) = 0$; otherwise, we define the miRNA m_j as a known miRNA to d_i . Hence, according to premises that functionally similar miRNAs may regulate similar diseases and similar diseases tend to associate with functionally similar miRNAs, it is reasonable to assume that the miRNA m_j is related to d_i if m_j is a potential miRNA to d_i , m_x is a most related miRNA to m_j , and m_x is also a known miRNA to d_i at the same time. Thereafter, based on this assumption, for any given disease $d_i \in S_D$ and any given miRNA $m_j \in S_M$, we can define the weight between d_i and m_j as follows:

$$DMW(i,j) = \begin{cases} \frac{1}{\exp\left(\max_{x \in MRM(m_j)} SMM(j,x)\right)} & \text{if } DMM(i,j) = 0 \text{ and } DMM(i,x) \neq 0, \\ \frac{1}{\exp(DMM(i,j))} & \text{if } DMM(i,j) \neq 0, \\ 0 & \text{otherwise.} \end{cases} \quad (15)$$

Step 2: Secondly, according to Equation (10), for any two given diseases d_i and d_j that belong to S_D , we define the weight between d_i and d_j as follows:

$$DW(i,j) = \frac{1}{\exp(FDD(i,j))}. \quad (16)$$

From Equation (14), it is easy to know that the higher the semantic similarity between d_i and d_j is, the smaller the weight between d_i and d_j will be.

Step 3: Similarly, according to Equation (11), for any two given miRNAs m_i and m_j that belong to S_M , we define the weight between m_i and m_j as follows:

$$MW(i,j) = \frac{1}{\exp(FMM(i,j))} \quad (17)$$

From Equation (15), it is also easy to know that the higher the functionally similarity between m_i and m_j is, the smaller the weight between m_i and m_j will be.

Thereafter, based on above three steps, for $i \in [1, D + M]$ and $j \in [1, D + M]$, a weighted miRNA–disease interactive network can finally be constructed as follows:

$$GFW(i, j) = \begin{cases} DW(d_i, d_j), & \text{if } i \in [1, D] \text{ and } j \in [1, D], \\ DMW(d_i, m_j), & \text{if } i \in [1, D] \text{ and } j \in [D, D + M], \\ DMW(m_i, d_j), & \text{if } i \in [D, D + M] \text{ and } j \in [1, D], \\ MW(m_i, m_j), & \text{if } i \in [D, D + M] \text{ and } j \in [D, D + M]. \end{cases} \quad (18)$$

4.8. Calculation of the Shortest Path Based on the Weighted Interactive Network

For any two given nodes A and B in the weighted interactive network G , supposing that there is a path P consisting of n hops such as $P_0(=A), P_1, P_2, \dots, P_n(=B)$ from A to B in G , then we define the weights of path P as $\sum_{i=0}^{n-1} GFW(i, i + 1)$. In addition, among all the paths from A to B in G , a path from A to B with smallest weights is called the shortest path from A to B . Thereafter, it is reasonable to assume that, for any two given nodes A and B in the weighted interactive network G , the smaller the weight of the shortest path from A to B is, the more related to each other the nodes A and B will be. Thus, based on this assumption, for any two given nodes A and B in the weighted interactive network G , we can design an algorithm for searching the shortest path from A to B in G as follows:

Step 1: Initially, we define that $S = \{V_0\}$ is a set consisting of an arbitrary node V_0 in G , T is a set consisting of all nodes in G other than V_0 , and DD is a matrix defined as follows:

$$DD(i, j) = \begin{cases} GFW(i, j) & \text{if } GFW(i, j) > 0, \\ \infty & \text{if } GFW(i, j) = 0, \end{cases} \quad (19)$$

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

Step 2: Next, we select a node V_k from T randomly, if V_k satisfies that $V_k \notin S$ and the distance from V_k to S is smaller than the distance from any other node other than V_k in T to S . Here, we define the distance from a node x in G to a node set V in G as the smallest value of the distances between x and all nodes in V .

Step 3: Thereafter, if $DD(i, j) > DD(i, k) + DD(k, j)$, then we further modify $DD(i, j)$ to $DD(i, j) = DD(i, k) + DD(k, j)$ in the matrix DD .

Step 4: After repeating step 2 and step 3 until all nodes in G are included in S , then it is obvious that we can transfer the matrix DD to a $(D + M) \times (D + M)$ dimensional shortest path matrix (SPM).

4.9. Calculation of the Shortest Path Based on the Weighted Interactive Network

Considering the fact that known miRNA–disease associations are very sparse, for a specific disease d_i and a specific miRNA m_j , as illustrated in Figure 5, in this section, we adopt the concept of T most similar neighbors to estimate the association between d_i and m_j according to the following steps:

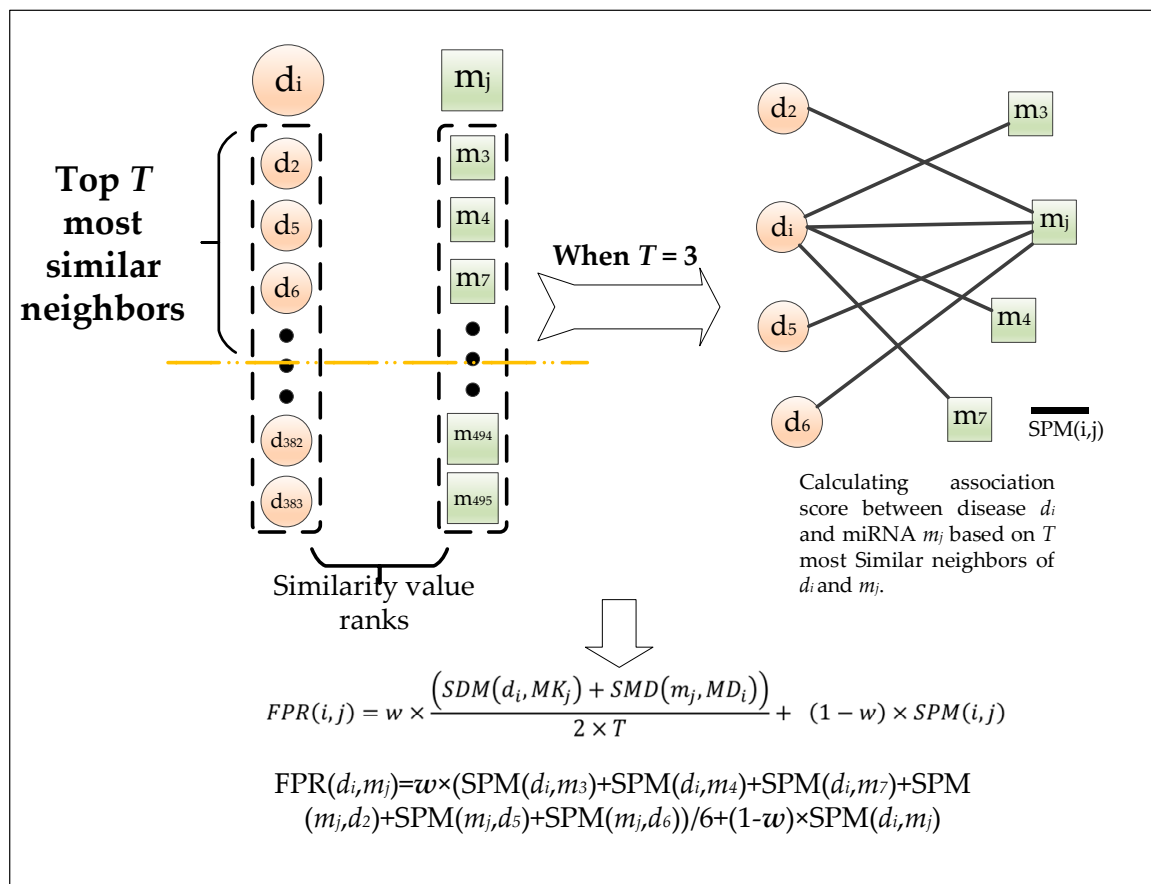


Figure 5. The process of predicting potential miRNA–disease associations based on the concept of *k* most similar neighbors.

Step 1: Firstly, for the disease d_i , let $DK_i = \{d_{i1}, d_{i2}, d_{i3} \dots, d_{iT}\}$ be a set consisting of the first T nodes in S_D after sorting the nodes in S_D by the disease integrated similarity between them with d_i in descending order, and, for the miRNA m_j , let $MK_j = \{m_{j1}, m_{j2}, m_{j3} \dots, m_{jT}\}$ be a set consisting of the first T nodes in S_M after sorting the nodes in S_M by the miRNA integrated similarity between them with m_j in descending order.

Step 2: Secondly, according to premises that functionally similar miRNAs may regulate similar diseases and similar diseases tend to associate with functionally similar miRNAs, we calculate the association between d_i and MK_j and the association between m_j and DK_i as follows:

$$SDM(d_i, MK_j) = \sum_{1 < q < T} SPM(d_i, m_{jq}) \tag{20}$$

$$SMD(m_j, MD_i) = \sum_{1 < q < T} SPM(m_j, d_{iq}) \tag{21}$$

Step 3: In order to optimize the prediction results, by integrating the above two associations and the matrix SPM , we can obtain our final prediction results as follows:

$$FPR(i, j) = w \times \frac{(SDM(d_i, MK_j) + SMD(m_j, MD_i))}{2 \times T} + (1 - w) \times SPM(i, j) \tag{22}$$

where w is a weight coefficient with a value from zero to one.

5. Conclusions

In this article, the effective predictive performance of WINMDA was mainly due to several reasons. Firstly, the semantic disease similarity, functional miRNA similarity, and Gaussian interaction profile kernel similarity were integrated. Secondly, we proposed a new method for calculating the semantic similarity of diseases. Thirdly, we constructed a weighted interactive network-based disease similarity, miRNA similarity, and known miRNA–disease associations. Fourthly, the concept of T most similar neighbours was introduced. Finally, an algorithm for searching the shortest path in the weighted interactive network was introduced. Furthermore, in future work, multiple heterogeneous biological data can be collected and pre-processed to be utilized in the weighted interactive network, thus improving the performance of prediction algorithms.

Supplementary Materials: Supplementary materials can be found at <http://www.mdpi.com/1422-0067/20/1/110/s1>.

Author Contributions: Conceptualization, H.Z. Data curation, H.Z. Formal analysis, L.K. and Q.Z. Funding acquisition, L.W. Investigation, L.K. Methodology, H.Z. Project administration, X.F., Q.Z., and L.W. Resources, X.F. Supervision, X.F. and L.W. Validation, L.K. Writing—original draft, H.Z. Writing—review and editing, L.K. and Q.Z.

Funding: This research was funded by the National Natural Science Foundation of China (No.61873221, No.61472282, No.61672447), the Natural Science Foundation of Hunan Province (No.2018JJ4058, No.2017JJ5036), and the CERNET Next Generation Internet Technology Innovation Project (No. NGII20160305, No. NGII20170109).

Acknowledgments: The authors thank the anonymous referees for suggestions that helped improve the paper substantially.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

miRNAs	MicroRNAs
WINMDA	Weighted interactive network for discovering potential miRNA–disease associations
CXCL8	C–X–C ligand 8
ROC	receiver operating characteristics
AUC	Area of ROC under the curve
LOOCV	Leave-one-out cross-validation
OC	Ovarian cancer
ncRNAs	Non-coding RNAs
MeSH	Medical Subject Headings
DAGs	Direct acyclic graphs
HMDD	Human microRNA Disease Database
LncRNA	long non-coding RNA
NEAT 1	nuclear paraspeckle assembly transcript 1
SNPs	single-nucleotide polymorphisms
dbDEMC	Differentially Expressed miRNAs in Human Cancers

References

- Chen, X.; Yan, G.Y. Novel human lncRNA–disease association inference based on lncRNA expression profiles. *Bioinformatics* **2013**, *29*, 2617–2624. [[CrossRef](#)] [[PubMed](#)]
- Consortium, E.P.; Birney, E.; Stamatoyannopoulos, J.A.; Dutta, A.; Guigó, R.; Gingeras, T.R.; Margulies, E.H.; Weng, Z.; Snyder, M.; Dermitzakis, E.T. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **2007**, *447*, 799–816. [[CrossRef](#)]
- Lander, E.S.; Linton, L.M.; Birren, B.; Nusbaum, C.; Zody, M.C.; Baldwin, J.; Devon, K.; Dewar, K.; Doyle, M.; FitzHugh, W. Initial sequencing and analysis of the human genome. *Nature* **2001**, *409*, 860–921. [[CrossRef](#)] [[PubMed](#)]
- Zou, Q.; Li, J.; Song, L.; Zeng, X.; Wang, G. Similarity computation strategies in the microRNA–disease network: A survey. *Brief. Funct. Genom.* **2016**, *15*, 55. [[CrossRef](#)] [[PubMed](#)]
- Ambros, V. The functions of animal microRNAs. *Nature* **2004**, *431*, 350–355. [[CrossRef](#)]
- Bartel, D.P. MicroRNAs: Target recognition and regulatory functions. *Cell* **2009**, *136*, 215–233. [[CrossRef](#)]

7. Khella, H.W.; Scorilas, A.; Mozes, R.; Mirham, L.; Lianidou, E.; Krylov, S.N.; Lee, J.Y.; Ordon, M.; Stewart, R.; Jewett, M.A. Low expression of miR-126 is a prognostic marker for metastatic clear cell renal cell carcinoma. *Am. J. Pathol.* **2015**, *185*, 693–703. [[CrossRef](#)]
8. Nan, D.; Wu, H.; Tao, T.; Peng, E. NEAT1 regulates cell proliferation and apoptosis of ovarian cancer by miR-34a-5p/BCL2. *Oncotargets Ther.* **2017**, *10*, 4905–4915. [[CrossRef](#)]
9. Xu, X.; Wu, X.; Jiang, Q.; Sun, Y.; Liu, H.; Chen, R.; Wu, S. Downregulation of microRNA-1 and microRNA-145 contributes synergistically to the development of colon cancer. *Int. J. Mol. Med.* **2015**, *36*, 1630–1638. [[CrossRef](#)]
10. Chen, B.; Duan, L.; Yin, G.; Jing, T.; Jiang, X. Simultaneously expressed miR-424 and miR-381 synergistically suppress the proliferation and survival of renal cancer cells—Cdc2 activity is up-regulated by targeting WEE1. *Clinics* **2013**, *68*, 825–833. [[CrossRef](#)]
11. Tang, W.; Wan, S.; Yang, Z.; Teschendorff, A.E.; Zou, Q. Tumor origin detection with tissue-specific miRNA and DNA methylation markers. *Bioinformatics* **2017**, *34*, 398–406. [[CrossRef](#)] [[PubMed](#)]
12. Zeng, X.; Liu, L.; Lü, L.; Zou, Q. Prediction of potential disease-associated microRNAs using structural perturbation method. *Bioinformatics* **2018**, *34*, 2425–2432. [[CrossRef](#)] [[PubMed](#)]
13. Ping, P.; Wang, L.; Kuang, L.; Ye, S.; Mfb, I.; Pei, T. A Novel method for lncRNA-disease association prediction based on an lncRNA-disease association network. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2018**, *1*. [[CrossRef](#)] [[PubMed](#)]
14. Yu, J.; Ping, P.; Wang, L.; Kuang, L.; Li, X.; Wu, Z. A Novel probability model for lncRNA-disease association prediction based on the Naïve Bayesian Classifier. *Genes* **2018**, *9*, 345. [[CrossRef](#)] [[PubMed](#)]
15. Kozomara, A.; Griffiths-Jones, S. miRBase: Annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* **2014**, *42*, D68. [[CrossRef](#)] [[PubMed](#)]
16. Yang, Z.; Ren, F.; Liu, C.; He, S.; Sun, G.; Gao, Q.; Yao, L.; Zhang, Y.; Miao, R.; Cao, Y. dbDEMC: A database of differentially expressed miRNAs in human cancers. *BMC Genom.* **2010**, *11*, S5. [[CrossRef](#)] [[PubMed](#)]
17. Li, Y.; Qiu, C.; Tu, J.; Geng, B.; Yang, J.; Jiang, T.; Cui, Q. HMDD v2.0: A database for experimentally supported human microRNA and disease associations. *Nucleic Acids Res.* **2014**, *42*, D1070. [[CrossRef](#)] [[PubMed](#)]
18. Jiang, Q.; Wang, Y.; Hao, Y.; Juan, L.; Teng, M.; Zhang, X.; Li, M.; Wang, G.; Liu, Y. miR2Disease: A manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res.* **2009**, *37*, D98–D104. [[CrossRef](#)]
19. Chen, X.; Wang, L.; Qu, J.; Guan, N.; Li, J. Predicting miRNA–disease association based on inductive matrix completion. *Bioinformatics* **2018**, *34*, 4256–4265. [[CrossRef](#)]
20. Chen, X.; Xie, D.; Zhao, Q.; You, Z.H. MicroRNAs and complex diseases: From experimental results to computational models. *Brief. Bioinform.* **2017**, *18*, 558. [[CrossRef](#)]
21. Chen, X.; Zhou, Z. ELLPMDA: Ensemble learning and link prediction for miRNA-disease association prediction. *RNA Biol.* **2018**, *15*, 807–818. [[CrossRef](#)] [[PubMed](#)]
22. Chen, X.; Zhang, D.H.; You, Z.H. A heterogeneous label propagation approach to explore the potential associations between miRNA and disease. *J. Transl. Med.* **2018**, *16*, 348. [[CrossRef](#)] [[PubMed](#)]
23. Chen, X.; Wang, C.C.; Yin, J.; You, Z.H. Novel human miRNA-disease association inference based on random forest. *Mol. Ther. Nucleic Acids* **2018**, *13*, 568–579. [[CrossRef](#)] [[PubMed](#)]
24. Chen, X.; Cheng, J.Y.; Yin, J. Predicting microRNA-disease associations using bipartite local models and hubness-aware regression. *RNA Biol.* **2018**, *15*, 1192–1205. [[CrossRef](#)]
25. Chen, X.; Liu, M.X.; Yan, G.Y. RWRMDA: Predicting novel human microRNA-disease associations. *Mol. Biosyst.* **2012**, *8*, 2792–2798. [[CrossRef](#)]
26. Chen, X.; Yan, G.Y. Semi-supervised learning for potential human microRNA-disease associations inference. *Sci. Rep.* **2014**, *4*, 5501. [[CrossRef](#)]
27. Chen, X.; Yan, C.C.; Zhang, X.; You, Z.H.; Deng, L.; Liu, Y.; Zhang, Y.; Dai, Q. WBSMDA: Within and between score for miRNA-disease association prediction. *Sci. Rep.* **2016**, *6*, 21106. [[CrossRef](#)]
28. Zou, Q.; Li, J.; Hong, Q.; Lin, Z.; Wu, Y.; Shi, H.; Ying, J. Prediction of microRNA-disease associations based on social network analysis methods. *Biomed. Res. Int.* **2015**, *2015*, 810514. [[CrossRef](#)]
29. You, Z.H.; Huang, Z.A.; Zhu, Z.; Yan, G.Y.; Li, Z.W.; Wen, Z.; Chen, X. PBMDA: A novel and effective path-based computational model for miRNA-disease association prediction. *PLoS Comput. Biol.* **2017**, *13*, e1005455. [[CrossRef](#)]

30. Chen, X.; Huang, L. LRSSLMDA: Laplacian regularized sparse subspace learning for miRNA-disease association prediction. *PLoS Comput. Biol.* **2017**, *13*, e1005912. [[CrossRef](#)]
31. Chen, X.; Yin, J.; Qu, J.; Huang, L. MDHGI: Matrix decomposition and heterogeneous graph inference for miRNA-disease association prediction. *PLoS Comput. Biol.* **2018**, *14*, e1006418. [[CrossRef](#)] [[PubMed](#)]
32. Zhao, H.; Kuang, L.; Wang, L.; Ping, P.; Xuan, Z.; Pei, T.; Wu, Z. Prediction of microRNA-disease associations based on distance correlation set. *BMC Bioinform.* **2018**, *19*, 141. [[CrossRef](#)] [[PubMed](#)]
33. Chen, X.; Huang, L.; Xie, D.; Zhao, Q. EGBMMDA: Extreme gradient boosting machine for miRNA-disease association prediction. *Cell Death Dis.* **2018**, *9*, 3. [[CrossRef](#)] [[PubMed](#)]
34. Chen, X.; Xie, D.; Wang, L.; Zhao, Q.; You, Z.H.; Liu, H. BNPMDA: Bipartite network projection for miRNA-disease association prediction. *Bioinformatics* **2018**, *34*, 3178–3186. [[CrossRef](#)] [[PubMed](#)]
35. Zhang, H.D.; Jiang, L.H.; Sun, D.W.; Li, J.; Ji, Z.L. The role of mir-130a in cancer. *Breast Cancer* **2017**, *24*, 521–527. [[CrossRef](#)] [[PubMed](#)]
36. Depke, J.L.; Onitilo, A.A. Coalition building and the intervention wheel to address breast cancer screening in hmong women. *Clin. Med. Res.* **2011**, *9*, 1–6. [[CrossRef](#)]
37. Finley, J.W.; Davis, C.D.; Feng, Y. Selenium from high selenium broccoli protects rats from colon cancer. *J. Nutr.* **2000**, *130*, 2384. [[CrossRef](#)]
38. Bakitas, M.; Ahles, T.A.; Skalla, K.; Brokaw, F.C.; Byock, I.; Hanscom, B.; Lyons, K.D.; Hegel, M.T. Proxy perspectives about end-of-life care for Person's with cancer. *Cancer* **2010**, *112*, 1854–1861. [[CrossRef](#)]
39. Heath, D.D.; Devlin, R.H.; Heath, J.W.; Iwama, G.K. Genetic, environmental and interaction effects on the incidence of jacking in *Oncorhynchus tshawytscha* (chinook salmon). *Heredity* **1994**, *72*, 146–154. [[CrossRef](#)]
40. Hollis, M.; Nair, K.; Vyas, A.; Chaturvedi, L.S.; Gambhir, S.; Vyas, D. MicroRNAs potential utility in colon cancer: Early detection, prognosis, and chemosensitivity. *World J. Gastroenterol.* **2015**, *21*, 8284–8292. [[CrossRef](#)]
41. Gomes, S.E.; Simões, A.E.S.; Pereira, D.M.; Rui, E.C.; Rodrigues, C.M.P.; Borralho, P.M. miR-143 or miR-145 overexpression increases cetuximab-mediated antibody-dependent cellular cytotoxicity in human colon cancer cells. *Oncotarget* **2016**, *7*, 9368–9387. [[CrossRef](#)] [[PubMed](#)]
42. Signs, S.A.; Fisher, R.C.; Tran, U.; Chakrabarti, S.; Sarvestani, S.K.; Xiang, S.; Liska, D.; Roche, V.; Lai, W.; Gittleman, H.R. Stromal miR-20a controls paracrine CXCL8 secretion in colitis and colon cancer. *Oncotarget* **2018**, *9*, 13048–13059. [[CrossRef](#)] [[PubMed](#)]
43. Jiang, H.; Ge, F.; Hu, B.; Wu, L.; Yang, H.; Wang, H. rs35301225 polymorphism in miR-34a promotes development of human colon cancer by deregulation of 3'UTR in E2F1 in Chinese population. *Cancer Cell Int.* **2017**, *17*, 39. [[CrossRef](#)] [[PubMed](#)]
44. Ding, S.; Eric, B.R.; Chen, Y.; Scull, B.; Kay, L.P.; Morgan, D. Molecular Imaging of Gastric Neoplasia with Near Infrared Fluorescent (NIRF) Activatable Probes. *Mol. Imaging* **2012**, *11*, 507–515. [[CrossRef](#)] [[PubMed](#)]
45. Zhou, L.; Zhao, X.; Han, Y.; Lu, Y.; Shang, Y.; Liu, C.; Li, T.; Jin, Z.; Fan, D.; Wu, K. Regulation of UHRF1 by miR-146a/b modulates gastric cancer invasion and metastasis. *FASEB J.* **2013**, *27*, 4929–4939. [[CrossRef](#)] [[PubMed](#)]
46. Zhuang, M.; Shi, Q.; Zhang, X.; Ding, Y.; Shan, L.; Shan, X.; Qian, J.; Zhou, X.; Huang, Z.; Zhu, W. Involvement of miR-143 in cisplatin resistance of gastric cancer cells via targeting IGF1R and BCL2. *Tumour Biol.* **2015**, *36*, 2737–2745. [[CrossRef](#)] [[PubMed](#)]
47. Marech, I.; Vacca, A.; Ranieri, G.; Gnoni, A.; Dammacco, F. Novel strategies in the treatment of castration-resistant prostate cancer (Review). *Int. J. Oncol.* **2012**, *40*, 1313–1320. [[CrossRef](#)]
48. Salinas, C.A.; Tsodikov, A.; Ishakhoward, M.; Cooney, K.A. Prostate Cancer in Young Men: An Important Clinical Entity. *Nat. Rev. Urol.* **2014**, *11*, 317–323. [[CrossRef](#)]
49. Chu, H.; Zhong, D.; Tang, J.; Li, J.; Xue, Y.; Tong, N.; Qin, C.; Yin, C.; Zhang, Z.; Wang, M. A functional variant in miR-143 promoter contributes to prostate cancer risk. *Arch. Toxicol.* **2016**, *90*, 403–414. [[CrossRef](#)]
50. Wang, D.; Lu, G.; Shao, Y.; Xu, D. MiR-182 promotes prostate cancer progression through activating Wnt/ β -catenin signal pathway. *Biomed. Pharmacother.* **2018**, *99*, 334–339. [[CrossRef](#)]
51. Taddei, M.L.; Cavallini, L.; Comito, G.; Giannoni, E.; Folini, M.; Marini, A.; Gandellini, P.; Morandi, A.; Pintus, G.; Raspollini, M.R. Senescent stroma promotes prostate cancer progression: The role of miR-210. *Mol. Oncol.* **2014**, *8*, 1729–1746. [[CrossRef](#)] [[PubMed](#)]

52. Wang, D.; Wang, J.; Lu, M.; Song, F.; Cui, Q. Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases. *Bioinformatics* **2010**, *26*, 1644–1650. [[CrossRef](#)] [[PubMed](#)]
53. Chen, X.; You, Z.H.; Yan, G.Y.; Gong, D.W. IRWRLDA: Improved random walk with restart for lncRNA-disease association prediction. *Oncotarget* **2016**, *7*, 57919–57931. [[CrossRef](#)] [[PubMed](#)]
54. Vanunu, O.; Magger, O.; Ruppin, E.; Shlomi, T.; Sharan, R. Associating genes and protein complexes with disease via network propagation. *PLoS Comput. Biol.* **2010**, *6*, e1000641. [[CrossRef](#)] [[PubMed](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).