

Article

A Semi-Supervised Learning Algorithm for Predicting Four Types MiRNA-Disease Associations by Mutual Information in a Heterogeneous Network

Xiaotian Zhang, Jian Yin * and Xu Zhang

School of Mechanical, Electrical and Information Engineering, Shandong University, Weihai 264209, China; zhangxiaotian@mail.sdu.edu.cn (Xiaotian.Z.); yunkaiyibi@163.com (Xu.Z.)

* Correspondence: yinjian@sdu.edu.cn; Tel.: +86-186-6313-0930

Received: 31 December 2017; Accepted: 22 February 2018; Published: 2 March 2018

Abstract: Increasing evidence suggests that dysregulation of microRNAs (miRNAs) may lead to a variety of diseases. Therefore, identifying disease-related miRNAs is a crucial problem. Currently, many computational approaches have been proposed to predict binary miRNA-disease associations. In this study, in order to predict underlying miRNA-disease association types, a semi-supervised model called the network-based label propagation algorithm is proposed to infer multiple types of miRNA-disease associations (NLPMMDA) by mutual information derived from the heterogeneous network. The NLPMMDA method integrates disease semantic similarity, miRNA functional similarity, and Gaussian interaction profile kernel similarity information of miRNAs and diseases to construct a heterogeneous network. NLPMMDA is a semi-supervised model which does not require verified negative samples. Leave-one-out cross validation (LOOCV) was implemented for four known types of miRNA-disease associations and demonstrated the reliable performance of our method. Moreover, case studies of lung cancer and breast cancer confirmed effective performance of NLPMMDA to predict novel miRNA-disease associations and their association types.

Keywords: multiple type miRNA-disease association prediction; semi-supervised learning; network similarity; label propagation algorithm

1. Introduction

MicroRNAs (miRNAs) are small endogenous non-coding RNAs that mainly regulate gene expression at the post-transcriptional level [1–3]. They are evolutionarily conserved and play a regulatory role by base pairing with messenger RNAs (mRNAs), resulting in mRNA degradation or translation inhibition [2,4,5]. Increasing evidence suggests that miRNAs are involved in a variety of critical biological processes, such as development, differentiation, apoptosis and metabolism [2]. Since the discovery of lin-4 and let-7 [6,7], many researchers have focused on the study of miRNAs, and numerous miRNAs have been identified. Furthermore, a great deal of databases have been established to provide information on miRNAs, such as the Human microRNA Disease Database (HMDD) [8], miR2Disease [9], database of Differentially Expressed miRNAs in human Cancers (dbDEMC) [10] and so on. It has been demonstrated that dysregulation of miRNAs may lead to a variety of diseases [11–13]. For example, miR-21 can target the *MAP2K3* gene directly during the carcinogenesis of hepatocellular carcinoma, resulting in expression inhibition of *MAP2K3* [14]. This also indicates that miRNAs can serve as efficient biomarkers for disease detection, diagnosis and prognosis [15]. Therefore, identifying disease-related miRNAs is a crucial problem.

During the past few decades, various disease-related miRNAs have been identified by several experimental methods. However, with the increasing of new miRNAs and other biological information,

these experimental methods face severe challenges, such as long experimental periods and expensive equipment. Thus, many computational approaches have been proposed to predict miRNA-disease associations [16]. The goal of computational approaches is to reduce the number of candidate miRNAs. Based on abundant biological information, several network-based methods are proposed to infer the relationships between miRNAs and diseases. The key of network-based methods is to calculate similarity scores among miRNAs and diseases over networks. It is well known that miRNAs with similar functions tend to be associated with similar diseases and vice versa [16,17]. Based on this assumption, Wang et al. [18] provided a method to infer human miRNA functional similarity (MISIM) by measuring semantic similarity of diseases which associated with miRNAs. Furthermore, they constructed a miRNA functional network. On the basis of results studied in [18], Xuan et al. [19] calculated miRNA functional similarity by integrating information of disease terms and similarity of disease phenotype. A new method was proposed to predict disease-related miRNAs, which is based on weighted k of most similar neighbors. The weights were assigned by miRNA family or cluster information. This method exhibited a good performance, but it was not applicable to diseases without any known associated miRNAs. Therefore, Chen and Zhang [20] adopted a network-consistency-based inference method to predict potential disease-related miRNAs based on the data derived from the miRNA functional similarity network, disease phenotype similarity network, and known miRNA-disease interaction network, which can be applied to isolated diseases without any related miRNAs. Nevertheless, the performance is not particularly satisfactory. Recently, Gu et al. [21] proposed a global and effective method to infer the associations between miRNAs and diseases, which is called network consistency projection for miRNA-disease associations (NCPMDA). NCPMDA is a non-parametric approach and takes full advantage of various molecular data, including miRNA functional similarity network, disease semantic similarity network, validated known miRNA-disease associations and miRNA family information. NCPMDA is applicable to isolated diseases and the predictive performance is superior to the previous method. The main problem of the above algorithms is that they all adopted similarity scores calculated by Wang et al. [18] that are estimated by miRNA-related diseases. However, the related disease information of many miRNAs remains largely unknown. In addition, these methods do not allow identification of the underlying proteins involved in miRNA-disease associations. In order to reveal the underlying proteins, Mørk et al. [22] developed a protein-driven method, miRPD, to infer miRNA-disease associations by using miRNA-protein associations and protein-disease associations. Without using any known information of miRNA-disease relationships, this method measured the associations of miRNAs and diseases via calculating association scores of miRNA-protein and protein-disease. The miRPD method attempted to reveal the underlying proteins involved in miRNA-disease associations and exhibited a reliable result.

Machine-learning-based algorithms are beneficial to improve the prediction performance, and up to now, several studies have proposed machine-learning-based models to predict potential miRNA-disease associations. For example, Jiang et al. [23] trained a support vector machine classifier by a feature vector to distinguish positive miRNA-disease associations from large-scale negative ones. Xu et al. [24] constructed a support vector machine classifier based on the features and changes in miRNA expression; then, the classifier was applied to an miRNA target-dysregulated network to infer new disease miRNAs. Generally, the common limitation of these methods is the selection of negative samples. It is known that there are no experimental validated negative associations between miRNAs and diseases, and miRNA-disease pairs selected from unknown miRNA-disease pairs might appear to be too biased. Considering this fact, Chen and Yan [25] proposed a semi-supervised method motivated by the framework of regularized least squares to infer potential miRNA-disease associations (RLSMDA). RLSMDA exhibited a reliable performance by leave-one-out cross validation (LOOCV) and case studies. In addition, it can work for diseases without any known related miRNAs. However, it might have biased the results, i.e., that they constructed disease similarity network and miRNA similarity network only by disease semantic similarity and miRNA functional similarity, respectively. Recently, based on

the hypothesis that distributional semantics can reveal information of relationships between miRNAs and diseases, Pasquier and Gardès [26] proposed a vector space model to discover new disease-miRNA associations. In this method, distributional information of miRNAs and diseases is represented in a high-dimensional vector space, which contains miRNA-disease associations, miRNA-related target mRNAs, family information of miRNAs, and genomic location information of miRNAs and abstracts of associated studies. By reducing the dimensionality of this high-dimensional vector space to fewer dimensions, they calculated the cosine distance of two vectors to measure their correlations. This method makes full use of miRNA-related information and achieves a satisfactory performance.

All of the above computational methods have identified various novel miRNA-disease interactions, but the specific types of interaction have not been predicted. Thus, mechanisms underlying the miRNA-disease associations still cannot be fully understood. In recent years, investigating the role of miRNAs in pathogenesis of human diseases has become one of the hottest topics [8], especially for entries from circulating miRNAs, epigenetics, miRNA-target interactions and genetics, whose number recorded in the HMDD has increased remarkably. As for interactions between miRNAs and targeted genes, for example, miRTar [27], an integrated system, identifies miRNA-target interactions in various scenarios and analyzes miRNA-targeted genes in pathways. In order to improve the accuracy of miRNA-gene target interaction identification, Pio et al. [28] presented a semi-supervised ensemble-based classifier that combines the prediction scores returned by several base algorithms to infer miRNA-targeted genes. They also predicted miRNA regulatory networks by a bi-clustering algorithm, which analyzes miRNA-target interactions to obtain inference results. The predicted miRNA-target interactions and miRNA regulatory network are stored in Co-clustered miRNA Regulatory Networks (ComiRNet) database. All of these researchers have shown that the miRNA regulatory network is complicated. Therefore, Chen et al. [29] developed a method to predict multiple types of miRNA-disease associations by a restricted Boltzmann machine (RBM) model. They constructed RBMs for miRNAs based on the data derived from HMDD v2.0, which included four types of miRNA-disease associations. Based on a contrastive divergence (CD) algorithm, they trained the constructed RBMs by initially setting the visible layer and hidden layer to obtain parameters of the RBM model. Finally, novel disease-related miRNAs and their types of interaction can be predicted by the trained RBM model. Although this method builds on the first model to predict multiple types of miRNA-disease associations, it only takes advantage of the data of known four types of miRNA-disease associations and ignores the relationships of disease-disease pairs and miRNA-miRNA pairs. Besides, RBM is a deep learning model and its training is time-consuming. In addition, the interaction prediction method among other types of biological entities can provide constructive suggestions for us in miRNA-disease interaction inference. By comparing over thirty network inference methods, Marbach et al. [30] observed that community-based methods can result in a powerful and robust performance for gene regulatory network reconstruction across different gold standards datasets. Therefore, Ceci et al. [31] proposed a semi-supervised method to deal with the problem of gene network reconstruction based on a multi-view learning framework. After assigning labels and identifying the multiple views, the method builds a classifier for each view, and then combines the output results of views to obtain final results. By applying a clustering algorithm, such as principle components analysis (PCA) or *k*-means, the views can be automatically identified by the system. This algorithm resolves the low quality and small quantity problem of known gene-gene interaction data and combines advantages of existing methods to achieve a good performance.

In this paper, a semi-supervised model called network-based label propagation method for inferring multiple types of miRNA-disease associations (NLPMDA) is proposed by mutual information derived from the heterogeneous network. Label propagation is an efficient algorithm which can make full use of the information of labeled and unlabeled data and has been used in many studies [32–35]. A key of the NLPMDA method is to construct a heterogeneous network. Firstly, a disease similarity homo-network is established by disease semantic similarity and Gaussian interaction profile kernel similarity. Secondly, a miRNA similarity homo-network is constructed

in a similar way, which combines miRNA functional similarity and Gaussian interaction profile kernel similarity. Thirdly, a multi-type miRNA-disease association hetero-network is established by validating four types of miRNA-disease associations. Then NLPMMDA performs label propagation in each homo-network. The homo-networks are used to capture cluster structure among diseases and miRNAs, and the hetero-network is used to capture mutual information of miRNA and disease pairs. Finally, final label scores of miRNA-disease pairs under four types can be calculated by propagating information on the heterogeneous network. The results of LOOCV and case studies demonstrated the reliable performance of NLPMMDA.

2. Materials and Methods

2.1. Data Preparation

In this paper, four types of human miRNA-disease association data were retrieved from HMDD [36]. In the updated database, human miRNA-disease data were annotated in four types, including entries from miRNA-target interactions, circulation samples, epigenetics and genetics [8]. After mapping the different miRNA precursors to mature miRNAs, the repeating miRNA-disease associations were removed. Finally, 682 miRNA-disease association data were obtained from miRNA-target interactions, 443 entries from circulations, 199 entries from epigenetics and 356 entries from genetics. All of these 1680 miRNA-disease associations are involved in 324 miRNAs and 171 diseases. These four types of miRNA-disease associations were used to construct a multi-type miRNA-disease association hetero-network which can offer the mutual interaction information. Besides, these four types of miRNA-disease associations are used as the gold standard dataset to evaluate the performance of our algorithm.

2.2. Construct Disease Similarity Homo-Network

The relationship of diseases can be represented by a directed acyclic graph (DAG) according to the disease classification system in the Medical Subject Headings (MeSH) database, in which nodes represent diseases and links represent the relationship of two diseases. For instance, a DAG of a disease d_i can be represented as $DAG(d_i) = (d_i, V(d_i), E(d_i))$, where $V(d_i)$ represents the vertices set of all ancestor diseases of d_i and disease d_i itself, and $E(d_i)$ represents the edges set of corresponding links. According to the algorithm proposed in [18], semantic similarity value SS of d_i and d_j can be calculated by:

$$SS(d_i, d_j) = \frac{\sum_{d \in V(d_i) \cap V(d_j)} (D_{d_i}(d) + D_{d_j}(d))}{SV(d_i) + SV(d_j)}, \quad (1)$$

where $D_{d_i}(d)$ is the contribution of disease d to the semantic value of disease d_i , the contribution of disease d_i itself to its own semantic value is defined as 1 and the contribution of other diseases is defined as $\max\{\Delta * D_{d_i}(d') | d' \in \text{children node of } d\}$. Here, Δ is the semantic contribution factor to distinguish the different semantic contribution values of disease d in different layers of $DAG(d_i)$; $SV(d_i)$ is the semantic value of disease d_i , which can be defined as $\sum_{d \in V(d_i)} D_{d_i}(d)$.

Gaussian interaction profile kernel similarity for diseases can be calculated by Gaussian kernel [37]. The miRNA interaction profile of a disease d_i is defined as $DIP(d_i)$, which is a binary vector to represent whether the disease d_i interacts with every miRNA in the multi-type miRNA-disease association hetero-network. Thus, the Gaussian interaction profile kernel similarity GS_d of disease d_i and disease d_j is defined as:

$$GS_d(d_i, d_j) = \exp(-\gamma_d \|DIP(d_i) - DIP(d_j)\|^2), \quad (2)$$

where γ_d is a parameter used to control the kernel bandwidth, which is set as $1/(\sum_{i=1}^{n_d} DIP(d_i)^2 / n_d)$. Here, n_d is the total number of diseases.

By integrating the disease semantic similarity matrix and Gaussian interaction profile kernel similarity matrix for diseases, disease similarity matrix S_d of disease similarity homo-network can be obtained as Equation (3).

$$S_d(d_i, d_j) = \begin{cases} SS(d_i, d_j), & d_i \text{ and } d_j \text{ have semantic similarity,} \\ GS(d_i, d_j), & \text{otherwise.} \end{cases} \quad (3)$$

In the disease similarity homo-network, the transition probability matrix is defined as:

$$P_d = D_d^{-\frac{1}{2}} S_d D_d^{-\frac{1}{2}}, \quad (4)$$

where D_d is a diagonal matrix and $D_d(i, i) = \sum_{j \in N_d} S_d(i, j)$, and N_d is the neighboring nodes set of disease d .

2.3. Construction of the miRNA Similarity Homo-Network

Similar to the construction of the disease similarity homo-network, the miRNA similarity homo-network is constructed based on miRNA functional similarity and Gaussian interaction profile kernel similarity. MiRNA functional similarity was calculated in a previous study [18]. The miRNA functional similarity value of miRNA m_i and m_j can be represented by $MFS(m_i, m_j)$. In order to reveal associations of miRNAs and diseases under different types, $MFS(m_i, m_j)$ is simply extended to multiple types of miRNA functional similarity matrix $MMFS(m_i, m_j, k)$, it is defined as:

$$MMFS(m_i, m_j, k) = MFS(m_{\lfloor i/n_k \rfloor}, m_{\lfloor j/n_k \rfloor}), \quad (5)$$

where k is the specific type, n_k is the total number of types.

The Gaussian interaction profile kernel similarity matrix for miRNAs can be calculated by:

$$GS_{m,k}(m_i, m_j) = \exp(-\gamma_{m,k} \|MIP_k(m_i) - MIP_k(m_j)\|^2), \quad (6)$$

where $MIP_k(m_i)$ is a binary vector which can represent relationships of miRNA m_i and the whole diseases under type k . $\gamma_{m,k}$ is a parameter used to control the kernel bandwidth which is set as $1/(\sum_{i=1}^{n_m} MIP_k(m_i)^2 / n_m)$. Here, n_m is the number of miRNAs.

The integrated miRNA similarity homo-network is constructed:

$$S_{m,k}(m_i, m_j) = \begin{cases} MMFS(m_i, m_j, k), & m_i \text{ and } m_j \text{ have functional similarity,} \\ GS_{m,k}(m_i, m_j), & \text{otherwise.} \end{cases} \quad (7)$$

In the miRNA similarity homo-network, the transition probability matrix is defined as:

$$P_{m,k} = D_{m,k}^{-\frac{1}{2}} S_{m,k} D_{m,k}^{-\frac{1}{2}}, \quad (8)$$

where $D_{m,k}$ is a diagonal matrix and $D_{m,k}(i, i) = \sum_{j,k \in N_m} S_{m,k}(i, j, k)$, and N_m is the neighboring nodes set of the miRNA m in miRNA homo-network.

2.4. Construction of the Multi-Type miRNA-Disease Association Hetero-Network

The multi-type miRNA-disease association hetero-network shows the relationships between miRNAs and diseases extracted from HMDD, including four types of human miRNA-disease association data. Figure 1 shows an example of the heterogeneous network, which contains four diseases and five miRNAs. The edges of multi-type miRNA-disease association hetero-network are created by four known types of miRNA-disease associations, and there are four edges between a disease and a miRNA at most. The edge vector $E_{ij} = \{e_k\}$ is used to represent the edges between

disease d_i and miRNA m_j , where $e_k = 1$ if d_i and m_j has an association of type k , and $e_k = 0$ otherwise. For example, if there are three association types between d_3 and m_2 , then the edge vector is $E_{32} = [1, 1, 1, 0]$. Based on the edge vectors, the adjacency matrix of multi-type miRNA-disease association hetero-network can be created. If disease d_i and miRNA m_j have confirmed associations, then $A(d_i, m_j) = E_{ij}$, where $i = 1, \dots, n_d, j = 1, \dots, n_m, n_d$ and n_m are the number of diseases and miRNAs, respectively.

Then, transition probability of miRNAs and diseases in hetero-network can be calculated by:

$$P_{d,m,k} = D_{d,m,k}^{-\frac{1}{2}} A D_{d,m,k}^{-\frac{1}{2}}, \tag{9}$$

where $D_{d,m,k}$ is a diagonal matrix and $D_{d,m,k}(i, i) = \sum_{j,k} A(i, j, k)$.

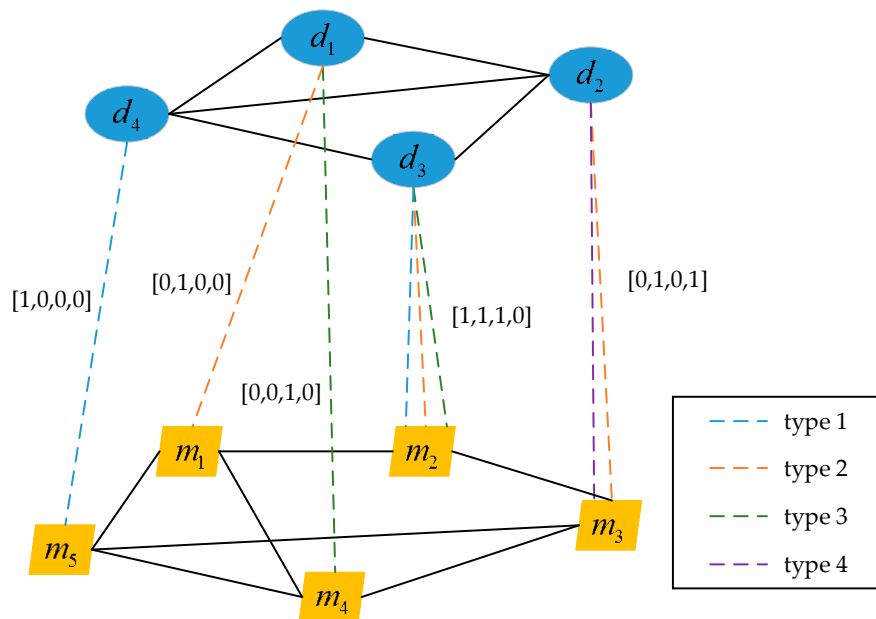


Figure 1. An example of the heterogeneous network composed of disease similarity homo-network, microRNAs (miRNA) similarity homo-network and multi-type miRNA-disease association hetero-network.

2.5. Network-Based Label Propagation Algorithm for Predicting Multiple miRNA-Disease Associations

Label propagation is a semi-supervised method. Its main purpose is to predict the labels of unlabeled data from both labeled and unlabeled data. A regularization framework for performing label propagation algorithm for a single network has been introduced and its convergence has been proved [35]. In this paper, label propagation is extended on a single network to our heterogeneous network, which is motivated by literature [38], and NLPMMDA is presented. Figure 2 shows the procedures of the NLPMMDA algorithm. The NLPMMDA method takes full advantage of mutual information in the heterogeneous network. Based on this method, novel disease-related miRNAs and the specific association types can be predicted.

The NLPMMDA algorithm can be described in detail as follows:

Step 1. Obtaining four types of miRNA-disease association data from HMDD and carrying out a data cleaning process.

Step 2. According to Sections 2.2–2.4, the heterogeneous network is constructed. In this study, the heterogeneous network $G = (V, E)$ is composed of the disease similarity homo-network $G_d = (V_d, E_d)$, miRNA similarity homo-network $G_{m,k} = (V_{m,k}, E_{m,k})$ and multi-type miRNA-disease association hetero-network $G_{d,m,k} = (V_d \cup V_{m,k}, E_{d,m,k})$.

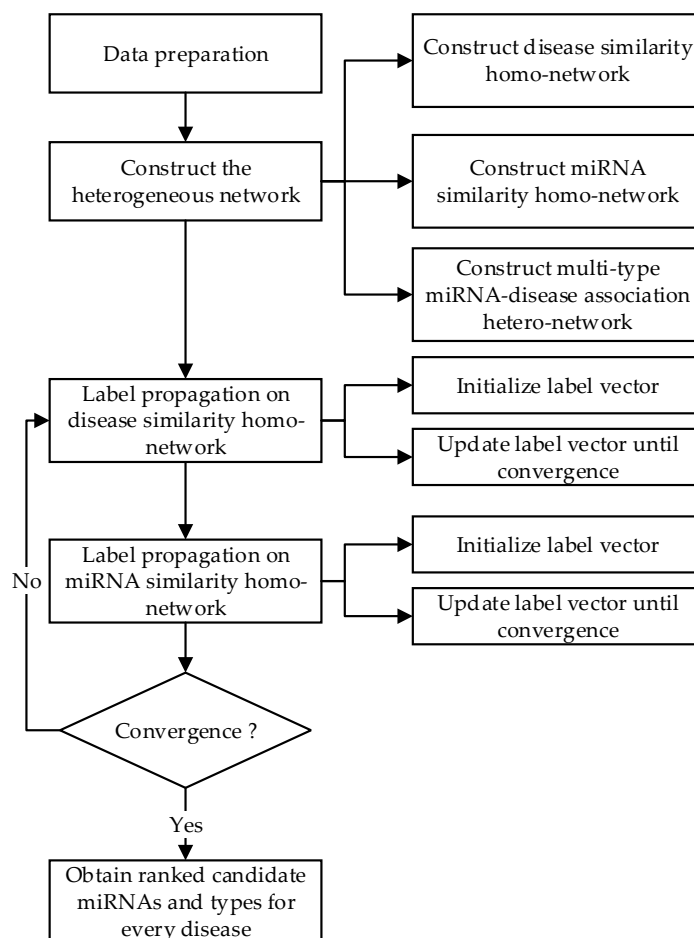


Figure 2. Flowchart of the Network-based Label Propagation Algorithm for Predicting Multiple miRNA-Disease Association (NLPMMDA).

Step 3. Performing network-based label propagation algorithm on the disease similarity homo-network. For a given query disease, the final label vector can be obtained by iteratively implementing Equation (10).

$$f_d^t = (1 - \lambda_d)f_d^0 + \lambda_d P_d f_d^{t-1}, \quad (10)$$

where P_d is the transition probability matrix calculated by Equation (4); f_d^{t-1} is a current label vector of diseases in which the i th element provides a current label score of disease d_i at time $t - 1$; f_d^t is the final label vector of diseases; f_d^0 is the initial label vector of disease nodes, and it can be obtained by Equation (11).

$$f_d^0 = \frac{1 - 2\lambda_d}{1 - \lambda_d} l_d^0 + \frac{\lambda_d}{1 - \lambda_d} P_{d,m,k} f_m, \quad (11)$$

where l_d^0 is the current label vector of diseases which is derived from miRNA-disease interaction hetero-network; f_m is the current label vector of miRNA nodes, λ_d is a diffusion parameter of disease similarity homo-network which specifies the relative amount of information from its neighbors and its initial label; $P_{d,m,k}$ is the transition probability matrix calculated by Equation (9). Finally, f_d^t converged to its limit f_d when $\|f_d^t - f_d^{t-1}\| < \sigma$, where σ is a threshold to control terminate iteration.

Step 4. Performing network-based label propagation algorithm on the miRNA similarity homo-network to obtain the final label vector according to Equation (12).

$$f_{m,k}^t = (1 - \lambda_m)f_{m,k}^0 + \lambda_m P_{m,k} f_{m,k}^{t-1}, \quad (12)$$

where $P_{m,k}$ is the transition probability matrix calculated by Equation (8); $f_{m,k}^{t-1}$ is the current label vector of miRNAs at time $t - 1$; $f_{m,k}^t$ is the final label vector of miRNAs; $f_{m,k}^0$ is the initial label vector of miRNAs in four types, which is calculated by Equation (13).

$$f_{m,k}^0 = \frac{1 - 2\lambda_m}{1 - \lambda_m} l_{m,k}^0 + \frac{\lambda_m}{1 - \lambda_m} P_{d,m,k} f_d, \quad (13)$$

where l_m^0 is the current label vector of miRNAs in which the j th element represents the current label score of miRNA m_j under type k ; f_d is the current label vector of diseases; λ_m is a diffusion parameter of miRNA similarity homo-network. Similarly, the condition of convergence is $\|f_{m,k}^t - f_{m,k}^{t-1}\| < \sigma$, where σ is a threshold to control terminate iteration.

Step 5. Sequentially implementing network-based label propagation in the disease similarity homo-network and miRNA similarity homo-network to update the final label vector f_m and f_d until both homo-networks converge. The condition of convergence is the same as mentioned above. Finally, for a given miRNA-disease pair, its final confidence label score in four types can be obtained. By ranking the label score in the final label vector, the top miRNAs are as considered as the most probable disease-related miRNAs and their type is considered as the most probable type.

3. Results

3.1. Performance Evaluation

In this study, to evaluate the performance of NLPMMDA, a LOOCV was implemented on four known and experimentally verified types of human miRNA-disease associations. Each known miRNA-disease association was left out in turn, and the remaining miRNA-disease associations were used as the labeled set. Then, the NLPMMDA method was implemented and the predictive scores of four types for each known miRNA-disease association were obtained. In addition, a receiver-operating characteristic (ROC) curve was drawn, which plots the true positive rate (TPR) versus the false positive rate (FPR) at different thresholds. The corresponding area under the ROC curve (AUC) was calculated to evaluate the predictive performance of the NLPMMDA method, where $AUC = 1$ means perfect performance and $AUC = 0.5$ means random performance. The ROC curve is typically used in binary classification problems to demonstrate the performance of a classifier. If a dataset only has positive and unlabeled samples, the ROC curve and AUC can be obtained by the ranked result of test samples. For example, in LOOCV, the test sample is ranked by the prediction scores of candidate miRNAs without confirmed association with currently investigated disease. In this paper, because the dataset can be divided into four classes, the output is operated by binarization and an ROC curve for each type is drawn. Finally, by considering each element of predictive scores as a binary prediction, the micro-average ROC curve was obtained. As can be seen in Figure 3, NLPMMDA obtained a reliable micro-average AUC value of 0.9739. The AUC value of four types of miRNA-disease associations is 0.9396, 0.9822, 0.9957 and 0.9813, respectively; type 1 represents entries from miRNA-target interactions, type 2 represents entries from circulation samples, type 3 represents entries from epigenetics and type 4 represents entries from genetics.

Besides, considering the limited number of known miRNA-disease associations, the area under the precision-recall (AUPR) curve is applied to further evaluate the performance of NLPMMDA. The precision-recall (PR) curve plots the relationship between precision and recall at different thresholds, where high precision is related to a low false positive rate, and high recall is related to a low false negative rate. Generally, an AUPR value closer to 1 means the performance is better. As shown in Figure 4, the micro-average AUPR value of NLPMMDA is 0.9323, and the AUPR value for every type is 0.9441, 0.9371, 0.9625 and 0.9225, respectively.

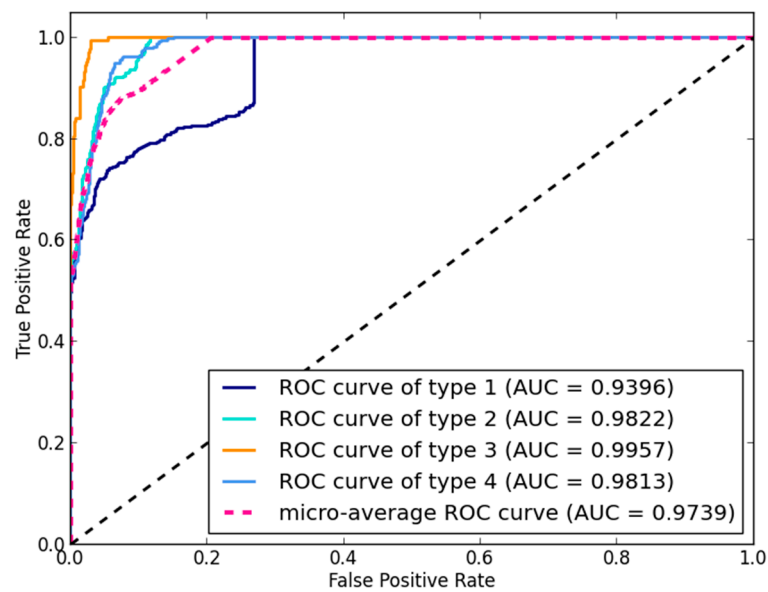


Figure 3. Receiver-operating characteristic (ROC) curve and area under the ROC curve (AUC) value of NLPMMDA based on leave-one-out cross validation (LOOCV). The micro-average AUC value of NLPMMDA is 0.9739. The AUC value of type 1, 2, 3, and 4 is 0.9396, 0.9822, 0.9957 and 0.9813, respectively.

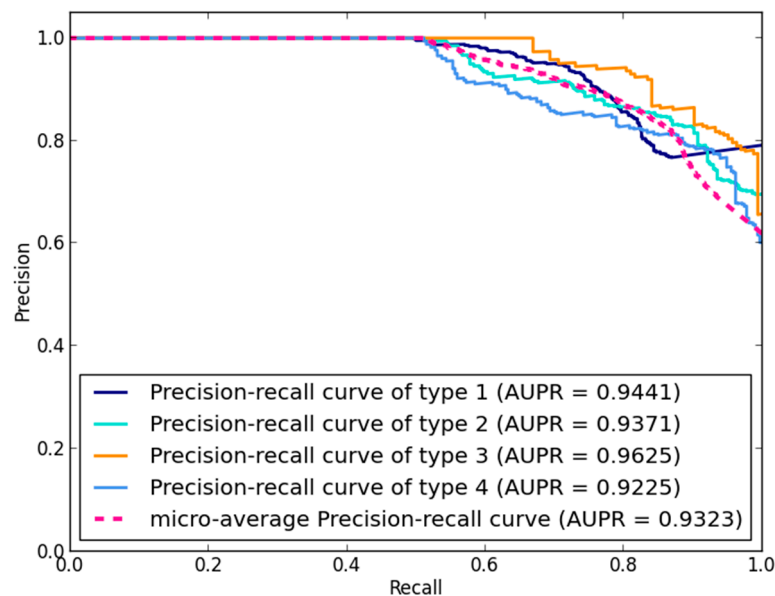


Figure 4. Precision-recall (PR) curve and area under the precision-recall (AUPR) value of NLPMMDA based on LOOCV. The micro-average AUPR value of NLPMMDA is 0.9323. The AUC value of type 1, 2, 3, 4 is 0.9441, 0.9371, 0.9625 and 0.9225, respectively.

3.2. Comparison with the Restricted Boltzmann Machine Model for Predicting Multiple Types of miRNA-Disease Associations Method

As far as we know, the restricted Boltzmann machine model for predicting multiple types of miRNA-disease associations (RBMMMDA) [29] is the first method to predict multiple types of miRNA-disease associations. It only makes use of known multiple types of miRNA-disease association data, and the AUC score of LOOCV is 0.8606. However, our method, NLPMMDA, integrates the

information of disease semantic similarity, Gaussian interaction profile kernel similarity for diseases, miRNA functionally similarity, Gaussian interaction profile kernel similarity for miRNAs and the known four types of miRNA-disease associations, obtaining a better performance. The micro-average AUC value of NLPMMDA is 0.9739. Considering the complex structure of the RBM model, it is difficult to combine the disease similarity information and miRNA similarity information in the RBM model. The performances of RBMMMDA and NLPMMDA can be seen in Table 1. In addition, the RBM model has various parameters and parameter selection problem is not solved well, thus the parameters of the RBM model are simply a used experience value. Parameters of the NLPMMDA method are selected by the performance of the experiment. Besides, training of the RBM model takes a long time. However, NLPMMDA is a semi-supervised method, and the execution time is short.

Table 1. Comparison with the restricted Boltzmann machine model for predicting multiple types of miRNA-disease associations (RBMMMDA) method.

Algorithms	RBMMMDA	NLPMMDA
AUC	0.8606	0.9739
Data	Known four types of miRNA-disease associations	Disease semantic similarity, miRNA functional similarity, Gaussian interaction profile kernel similarity and known four types of miRNA-disease associations
Application	Cannot be applied to isolated diseases	Cannot be applied to isolated diseases
Parameters	Use the previous value	Select by the performance of experiments
model	Supervised learning	Semi-supervised learning
Case study	Lung cancer: 33 of top 50	Lung cancer: 44 of top 50
	Breast cancer: 17 of top 50	Breast cancer: 37 of top 50

3.3. Effect of the Parameters

There are two parameters λ_d and λ_m in the NLPMMDA algorithm. λ_d is a diffusion parameter of disease similarity homo-network, which adjusts the relative amount of information from its initial label to its neighbors. λ_m is a diffusion parameter of miRNA similarity homo-network. In this paper, λ_d and λ_m are set to the same value. By selecting different λ_d and λ_m values (varying from 0.1 to 0.9 with scale 0.1), LOOCV is implemented to obtain the AUC score of the NLPMMDA method. The LOOCV results are shown in Table 2. As a result, the AUC value is almost equal in the range of $0.1 \leq \lambda_d \leq 0.4$ and $0.1 \leq \lambda_m \leq 0.4$, and AUC value is decreased in the range of $0.6 \leq \lambda_d \leq 0.9$ and $0.6 \leq \lambda_m \leq 0.9$. However, our predictive method has no predictive ability when λ_d and λ_m are equal to 0.5, which is a result of the approach of initialization in homo-networks. Therefore, in this study, $\lambda_d = 0.2$ and $\lambda_m = 0.2$ are selected to predict novel miRNA-disease association types by the NLPMMDA algorithm. The optimal values of parameters depend on the known miRNA-disease association dataset.

Table 2. Effect of the parameters.

λ_d	λ_m	AUC	AUPR
0.1	0.1	0.9738	0.9320
0.2	0.2	0.9739	0.9323
0.3	0.3	0.9738	0.9309
0.4	0.4	0.9720	0.9302
0.5	0.5	0.5	0.5
0.6	0.6	0.8173	0.6490
0.7	0.7	0.8076	0.6409
0.8	0.8	0.7900	0.6251
0.9	0.9	0.7559	0.5962

3.4. Case Studies of Lung Cancer and Breast Cancer

To further confirm the robustness of the NLPMMDA method, case studies of lung cancer and breast cancer were implemented to evaluate the ability of the NLPMMDA method for predicting multi-types of miRNA-disease associations. All known miRNA-disease associations under four types were assigned as labeled data, and unknown miRNA-disease pairs were used as unlabeled data. Then, based on labeled and unlabeled data, NLPMMDA can predict miRNA-disease relationships and their specific types. Prediction results were manually verified by online databases and recent literature. The top 50 potential miRNA-disease association types of lung cancer and breast cancer are listed in Tables 3 and 4, respectively, including disease-related miRNAs, miRNA-disease association types and evidences related to miRNA-disease pairs. The evidence is the PubMed Unique Identifier (PMID) of related literature. Due to the complexity of diseases and the associated miRNA roles, a predicted association type supported by three PubMed articles at least can be considered as a reliable association type.

The morbidity and mortality of lung cancer is high in both men and women, and lung cancer is the most common cause of cancer death worldwide [39]. Although various new therapeutics and strategies for detection and early diagnosis have progressed in lung cancer, its prognosis remains poor [40]. Recent studies demonstrated the important role of miRNAs in development and therapy response of lung cancer. In the labeled data, there are 52 miRNA-disease associations, which are classified as the miRNA-target type [41,42], circulating miRNA type [43,44], epigenetics type [45] and genetics type [46,47]. After implementing the NLPMMDA method on labeled and unlabeled data, scores of miRNA-disease pairs are predicted. As a result, among top 20 and top 50 candidates without relevance of known association types, 17 and 44 lung cancer-related miRNAs and their association types are supported by different evidence, respectively, and 25 predicted results are considered as reliable association types. As shown in Table 3, in the top 50 potential lung cancer-related miRNAs, miR-133a plays a tumor suppressor role in non-small cell lung cancer (NSCLC) by targeting *IGF-1R*, *TGFBR1* and *EGFR* [48]. Also, in NSCLC, miR-143 targets *ATG2B* and miR-34a targets *TGFβR2* to inhibit cell proliferation [49,50]; Besides, serum miR-126 and miR-21 levels can be used as novel biomarkers in non-small cell lung cancer development, metastasis and screening [51,52], and circulating miR-29a shows a highly prognostic signature in non-squamous NSCLC patients [53]. The single nucleotide polymorphisms rs2910164 of miR-146a are associated with the risk of NSCLC in the Chinese population, which can be regarded as the genetics type [54].

Table 3. Lung cancer-related candidate miRNAs and association types predicted by NLPMMDA.

miRNAs	Types	PMID	miRNAs	Types	PMID
hsa-mir-499a	genetics	unconfirmed	hsa-mir-19a	target	27588137;25604748;28592790
hsa-mir-146a	genetics	25154761;24144839;29127520	hsa-let-7f	target	29017393
hsa-mir-133a	target	24816813;22089643;25518741	hsa-mir-15a	target	25442346;24500260;25874488
hsa-mir-126	circulation	28253725;27093275;29266846	hsa-mir-206	target	26919096;26075299;25522678
hsa-mir-17	genetics	17384677	hsa-mir-16	genetics	unconfirmed
hsa-mir-21	circulation	25501703;25421010;29163821	hsa-mir-126	target	18602365;22510476;29277611
hsa-mir-143	target	25322940;25003638;24070896	hsa-mir-125b	target	28713974
hsa-mir-34a	target	25501507;25038915;24983493	hsa-mir-218	target	21159652;24247270;24705471
hsa-mir-20a	genetics	17384677	hsa-mir-17	circulation	23263848
hsa-mir-29a	circulation	24928469	hsa-let-7e	target	unconfirmed
hsa-mir-200c	target	24997798;24205206;23708087	hsa-mir-20a	target	24722426
hsa-mir-17	target	24755562;24722426;29289833	hsa-mir-219	target	28714014
hsa-mir-92a	genetics	unconfirmed	hsa-mir-222	target	21042732
hsa-mir-20a	circulation	25421010	hsa-mir-19b	target	28364280
hsa-mir-34a	epigenetics	18719384	hsa-mir-429	target	24866238;27602157
hsa-mir-34b	epigenetics	24130071;22047961;21383543	hsa-mir-223	circulation	28356944;25421010;29212284
hsa-mir-18a	genetics	unconfirmed	hsa-mir-18a	target	28471447
hsa-mir-200b	target	22139708;28731781;28615992	hsa-mir-122	circulation	24282590;25926378
hsa-mir-155	target	22027557;29260515;28939896	hsa-let-7a	target	21097396
hsa-mir-16	target	25435430;23954293;29138833	hsa-mir-15a	genetics	unconfirmed
hsa-mir-34c	epigenetics	24130071;22047961;21383543	hsa-mir-124	epigenetics	17308079
hsa-mir-221	target	18246122;21042732;19962668	hsa-mir-92a	target	23820254
hsa-mir-183	target	18840437;26951513;27593936	hsa-mir-133b	target	22883469;19654003;29328427
hsa-mir-214	target	28396596;26462018;28396596	hsa-mir-155	genetics	28225782
hsa-mir-146a	circulation	28678319;25755772;24531034	hsa-mir-203	target	25140799;24040137;28921827

PMID: PubMed Unique Identifier.

Based on annual statistical data, breast cancer is one of the most common types of cancer which mainly occurs in women [55]. Current studies demonstrated related death rates of breast cancer are still on the rise [56]. Besides, accumulating evidence shows that miRNAs play a vital role in breast cancer and can be used as diagnosis and therapeutic biomarkers for breast cancer patients. In our labeled data, there are 176 known miRNAs-disease associations which can be divided into four types according to evidence from literature. For example, serum miR-155 is up-regulated in breast cancer patients; thus, serum miR-155 is a potential biomarker to track breast cancer [57,58]. According to HMDD, the association between miR-155 and breast cancer is labeled as the circulation type [8]. The candidate miRNAs without known breast cancer-related miRNAs and their association types are predictive by the NLPMDA method. Among the top 20 and top 50 potential miRNAs, 17 and 37 miRNA-disease association types are confirmed by biological evidence, respectively, and 16 predicted results are considered as reliable association types. Table 4 shows the details. Hsa-miR-1 is a breast cancer-related miRNA in the HMDD database. However, their underlying association type is not clear. In our predictive result, the relationship between hsa-miR-1 and breast cancer is target type, which can be proved by various evidence. For example, as described in the result of Liu et al. [59], hsa-miR-1 can function as a tumor suppressor in breast cancer by targeting *K-RAS* and *MALAT1*. Also, *IMPDH1* and *NPEPL1* genes are identified as direct targets of miR-19a in breast cancer by a quantitative proteomic strategy [60]; miR-19b can promote metastasis of breast cancer by targeting *MYLIP* and its related cell adhesion molecules [61]; and miR-133a acts as a tumor suppressor in breast cancer by targeting *EGFR* [62]. Moreover, the plasma level of circulating miR-146a is involved in breast cancer biology and tumor progression [63]. In primary human breast cancer, hsa-miR-9 is affected by epigenetic inactivation because of aberrant hypermethylation [64].

Table 4. Breast cancer-related candidate miRNAs and association types predicted by NLPMDA.

miRNAs	Types	PMID	miRNAs	Types	PMID
hsa-mir-16	genetics	16754881;17012848	hsa-mir-127	target	24282530;24155205;25477702
hsa-mir-1	target	26275461;26926567;26497855	hsa-let-7i	target	24662829;21826373;
hsa-mir-126	circulation	28683441	hsa-let-7a	genetics	26681038
hsa-mir-19a	target	22952885;23831570;27596294	hsa-mir-106b	target	27519168;27325313;28518139
hsa-let-7a	target	24172884	hsa-mir-219	target	Unconfirmed
hsa-mir-19b	target	28969074;28731027;27602768	hsa-let-7f	genetics	23042301
hsa-mir-92a	genetics	Unconfirmed	hsa-mir-127	epigenetics	27998789
hsa-mir-223	circulation	Unconfirmed	hsa-mir-15b	target	25783158
hsa-mir-18a	target	19684618;25069832;21755340	hsa-mir-143	target	28746466;28559978;28588724;27121210
hsa-mir-29a	circulation	Unconfirmed	hsa-mir-19b	circulation	Unconfirmed
hsa-let-7c	target	25388283	hsa-mir-199a	circulation	26476723;25906045
hsa-mir-125b	genetics	19738052	hsa-let-7e	genetics	Unconfirmed
hsa-mir-133a	target	23786162;29207145;26107945	hsa-mir-145	circulation	23334650
hsa-mir-15a	target	27596816;27713175;28655885	hsa-mir-155	genetics	26095675
hsa-let-7d	target	22081076	hsa-let-7d	genetics	Unconfirmed
hsa-let-7f	target	22407818;25552929	hsa-mir-218	circulation	Unconfirmed
hsa-mir-29b	epigenetics	24297604	hsa-mir-221	circulation	25009660;22156446
hsa-mir-214	target	24577056;25738546;28071724	hsa-mir-146a	target	27175941;25596948;25712342
hsa-mir-9	epigenetics	26519551;17948228	hsa-mir-124	epigenetics	Unconfirmed
hsa-mir-146a	circulation	27197674;26033453;23898484	hsa-mir-19a	circulation	24938880;24416156
hsa-let-7e	target	Unconfirmed	hsa-let-7g	target	21868760
hsa-mir-18a	circulation	24694649;23705859;28109133	hsa-mir-106a	target	27325313
hsa-mir-25	target	25026296;29310680;28188287	hsa-mir-9	circulation	Unconfirmed
hsa-let-7b	target	21826373;24264599;23339187;22761738	hsa-mir-145	genetics	Unconfirmed
hsa-mir-92a	target	28881597;29162724;28881597	hsa-mir-19b	epigenetics	Unconfirmed

In conclusion, 44 and 37 out of the top 50 predictive lung cancer-related and breast cancer-related miRNAs and their specific association types are confirmed by experimental evidence, respectively. The results of case studies demonstrated the robustness of NLPMDA method.

3.5. Web Server for Network-Based Label Propagation Algorithm to Predicting Multiple miRNA-Disease Association Method

In this study, a web server was built to show the prediction results of the NLPMDA method, which is freely available at <http://39.107.230.144/NLPMDA>.

The web server enables the function of predicting four types of miRNA-disease associations based on the NLPMMDA algorithm. The final prediction result for a specific disease will be shown in a table, and the rank, miRNA name, association type and potential association probability will be included. The tables contain known verified related miRNAs and types for a disease, whose value of potential association probability is 1.0.

4. Discussion

Increasing evidence indicates the prominent role of miRNAs in the development of various diseases. Understanding the underlying mechanisms of miRNAs in diseases is becoming an urgent problem worldwide. In this study, a network-based label propagation algorithm is proposed to infer specific types of miRNA-disease associations, which integrated four types of known human miRNA-disease associations derived from HMDD. The NLPMMDA method constructed a heterogeneous network, in which a disease similarity homo-network is constructed by integrating disease semantic similarity information with Gaussian interaction profile kernel similarity information, and miRNA similarity homo-network is constructed by integrating miRNA functional similarity information with Gaussian interaction profile kernel similarity information. Besides, a multi-type miRNA-disease interaction hetero-network is constructed by four types of known miRNA-disease association data. In addition, the traditional label propagation algorithm is extended to the heterogeneous network and the strategy of label initialization is changed in the NLPMMDA method. The LOOCV result, case studies of lung cancer, and breast cancer demonstrate the reliable performance of the NLPMMDA method.

Compared with current computational methods which can predict multiple type miRNA-disease associations, the NLPMMDA method achieves a better performance because of several factors. Firstly, the network-based label propagation algorithm is a semi-supervised machine learning model. As we all know, one of the current difficulties of predictive models is the selection of negative samples. NLPMMDA does not require verified negative miRNA-disease associations. Secondly, transition probability among diseases and miRNAs under four types are calculated in the NLPMMDA method, which can capture the similarity information from neighboring nodes in homo-networks and improve the predictive function of the computational model. Thirdly, construction of heterogeneous network could offer mutual information between the miRNA similarity homo-network and disease homo-network. The label values of nodes in the homo-networks are initialized by their initial labels and neighbors from other homo-networks, which makes label confidence score more reliable. Although NLPMMDA exhibited highly reliable results, it still has some limitations. Transition probability scores among four types are simply calculated by miRNA functional similarity and Gaussian interaction profile kernel similarity, which may result in offset error. In addition, the NLPMMDA method is not applicable to diseases without any known associations of miRNAs. The different combination of diffusion parameters in homo-networks may improve the performance of the NLPMMDA method, which can be further studied in the future.

Acknowledgments: There is no source of funding for the study.

Author Contributions: Xiaotian Zhang and Jian Yin conceived and designed the experiments; Xiaotian Zhang performed the experiments; Xiaotian Zhang and Xu Zhang analyzed the experimental results; Xiaotian Zhang wrote the paper. All authors have read and approved the final manuscript.

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

References

1. Bartel, D.P. MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell* **2004**, *116*, 281–297. [[CrossRef](#)]
2. Bartel, D.P. MicroRNAs: Target recognition and regulatory functions. *Cell* **2009**, *136*, 215–233. [[CrossRef](#)] [[PubMed](#)]
3. Ambros, V. The functions of animal microRNAs. *Nature* **2004**, *431*, 350–355. [[CrossRef](#)] [[PubMed](#)]

4. Krol, J.; Loedige, I.; Filipowicz, W. The widespread regulation of microRNA biogenesis, function and decay. *Nat. Rev. Genet.* **2010**, *11*, 597–610. [[CrossRef](#)] [[PubMed](#)]
5. Cuperus, J.T.; Fahlgren, N.; Carrington, J.C. Evolution and functional diversification of miRNA genes. *Plant Cell* **2011**, *23*, 431–442. [[CrossRef](#)] [[PubMed](#)]
6. Lee, R.C.; Feinbaum, R.L.; Ambros, V. The *C. Elegans* heterochronic gene *lin-4* encodes small rnas with antisense complementarity to *lin-14*. *Cell* **1993**, *75*, 843–854. [[CrossRef](#)]
7. Reinhart, B.J.; Slack, F.J.; Basson, M.; Pasquinelli, A.E.; Bettinger, J.C.; Rougvie, A.E.; Horvitz, H.R.; Ruvkun, G. The 21-nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis Elegans*. *Nature* **2000**, *403*, 901–906. [[CrossRef](#)] [[PubMed](#)]
8. 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–D1074. [[CrossRef](#)] [[PubMed](#)]
9. 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](#)] [[PubMed](#)]
10. 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](#)]
11. Olson, E.N. MicroRNAs as therapeutic targets and biomarkers of cardiovascular disease. *Sci. Transl. Med.* **2014**, *6*, 239ps3. [[CrossRef](#)] [[PubMed](#)]
12. Kong, Y.W.; Ferland-McCollough, D.; Jackson, T.J.; Bushell, M. MicroRNAs in cancer management. *Lancet Oncol.* **2012**, *13*, e249–e258. [[CrossRef](#)]
13. Hatziapostolou, M.; Polytarchou, C.; Aggelidou, E.; Drakaki, A.; Poultsides, G.A.; Jaeger, S.A.; Ogata, H.; Karin, M.; Struhl, K.; Hadzopoulou-Cladaras, M. An HNF4 α -miRNA inflammatory feedback circuit regulates hepatocellular oncogenesis. *Cell* **2011**, *147*, 1233–1247. [[CrossRef](#)] [[PubMed](#)]
14. Xu, G.; Zhang, Y.; Wei, J.; Jia, W.; Ge, Z.; Zhang, Z.; Liu, X. MicroRNA-21 promotes hepatocellular carcinoma HepG2 cell proliferation through repression of mitogen-activated protein kinase-kinase 3. *BMC Cancer* **2013**, *13*, 469. [[CrossRef](#)] [[PubMed](#)]
15. Png, K.J.; Halberg, N.; Yoshida, M.; Tavazoie, S.F. A microRNA regulon that mediates endothelial recruitment and metastasis by cancer cells. *Nature* **2012**, *481*, 190–196. [[CrossRef](#)] [[PubMed](#)]
16. Zeng, X.; Zhang, X.; Zou, Q. Integrative approaches for predicting microRNA function and prioritizing disease-related microRNA using biological interaction networks. *Brief. Bioinform.* **2016**, *17*, 193–203. [[CrossRef](#)] [[PubMed](#)]
17. Jiang, Q.; Hao, Y.; Wang, G.; Juan, L.; Zhang, T.; Teng, M.; Liu, Y.; Wang, Y. Prioritization of disease microRNAs through a human phenome-microRNAome network. *BMC Syst. Biol.* **2010**, *4*, S2. [[CrossRef](#)] [[PubMed](#)]
18. Wang, D.; Wang, J.; Lu, M.; Song, F.; Cui, Q. Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases. *Bioinformatics* **2010**, *26*, 1644–1650. [[CrossRef](#)] [[PubMed](#)]
19. Xuan, P.; Han, K.; Guo, M.; Guo, Y.; Li, J.; Ding, J.; Liu, Y.; Dai, Q.; Li, J.; Teng, Z. Correction: Prediction of microRNAs associated with human diseases based on weighted *k* most similar neighbors. *PLoS ONE* **2013**, *8*, e70204. [[CrossRef](#)] [[PubMed](#)]
20. Chen, H.; Zhang, Z. Similarity-based methods for potential human microRNA-disease association prediction. *BMC Med. Genom.* **2013**, *6*, 1–9. [[CrossRef](#)] [[PubMed](#)]
21. Gu, C.; Bo, L.; Li, X.; Li, K. Network consistency projection for human miRNA-disease associations inference. *Sci. Rep.* **2016**, *6*, 36054. [[CrossRef](#)] [[PubMed](#)]
22. Mørk, S.; Pletscher-Frankild, S.; Caro, A.P.; Gorodkin, J.; Jensen, L.J. Protein-driven inference of miRNA–disease associations. *Bioinformatics* **2014**, *30*, 392–397. [[CrossRef](#)] [[PubMed](#)]
23. Jiang, Q.; Wang, G.; Jin, S.; Li, Y.; Wang, Y. Predicting human microRNA-disease associations based on support vector machine. *Int. J. Data Min. Bioinform.* **2013**, *8*, 282–293. [[CrossRef](#)] [[PubMed](#)]
24. Xu, J.; Li, C.-X.; Lv, J.-Y.; Li, Y.-S.; Xiao, Y.; Shao, T.-T.; Huo, X.; Li, X.; Zou, Y.; Han, Q.-L. Prioritizing candidate disease miRNAs by topological features in the miRNA target–dysregulated network: Case study of prostate cancer. *Mol. Cancer Ther.* **2011**, *10*, 1857–1866. [[CrossRef](#)] [[PubMed](#)]
25. Chen, X.; Yan, G.Y. Semi-supervised learning for potential human microRNA-disease associations inference. *Sci. Rep.* **2014**, *4*, 5501. [[CrossRef](#)] [[PubMed](#)]

26. Pasquier, C.; Gardès, J. Prediction of miRNA-disease associations with a vector space model. *Sci. Rep.* **2016**, *6*, 27036. [[CrossRef](#)] [[PubMed](#)]
27. Hsu, J.B.; Chiu, C.M.; Hsu, S.D.; Huang, W.Y.; Chien, C.H.; Lee, T.Y.; Huang, H.D. miRTar: An integrated system for identifying miRNA-target interactions in human. *BMC Bioinform.* **2011**, *12*, 300. [[CrossRef](#)] [[PubMed](#)]
28. Pio, G.; Ceci, M.; Malerba, D.; D'Elia, D. ComiRNet: A web-based system for the analysis of miRNA-gene regulatory networks. *BMC Bioinform.* **2015**, *16* (Suppl. S9), S7. [[CrossRef](#)] [[PubMed](#)]
29. Chen, X.; Yan, C.C.; Zhang, X.; Li, Z.; Deng, L.; Zhang, Y.; Dai, Q. RBMMMDA: Predicting multiple types of disease-microRNA associations. *Sci. Rep.* **2015**, *5*, 13877. [[CrossRef](#)] [[PubMed](#)]
30. Marbach, D.; Costello, J.C.; Küffner, R.; Vega, N.M.; Prill, R.J.; Camacho, D.M.; Allison, K.R.; Aderhold, A.; Allison, K.R.; Bonneau, R. Wisdom of crowds for robust gene network inference. *Nat. Methods* **2012**, *9*, 796–804. [[CrossRef](#)] [[PubMed](#)]
31. Ceci, M.; Pio, G.; Kuzmanovski, V.; Džeroski, S. Semi-supervised multi-view learning for gene network reconstruction. *PLoS ONE* **2015**, *10*, e0144031. [[CrossRef](#)] [[PubMed](#)]
32. Yan, X.-Y.; Zhang, S.-W.; Zhang, S.-Y. Prediction of drug–target interaction by label propagation with mutual interaction information derived from heterogeneous network. *Mol. BioSyst.* **2016**, *12*, 520–531. [[CrossRef](#)] [[PubMed](#)]
33. Ni, B.; Yan, S.; Kassim, A. Learning a propagable graph for semisupervised learning: Classification and regression. *IEEE Trans. Knowl. Data Eng.* **2012**, *24*, 114–126.
34. Huang, Y.-F.; Yeh, H.-Y.; Soo, V.-W. Inferring drug-disease associations from integration of chemical, genomic and phenotype data using network propagation. *BMC Med. Genom.* **2013**, *6*, S4. [[CrossRef](#)] [[PubMed](#)]
35. Zhou, D.; Bousquet, O.; Lal, T.N.; Weston, J.; Olkoph, B.S. Learning with local and global consistency. *Adv. Neural Inf. Process. Syst.* **2004**, *16*, 321–328.
36. Lu, M.; Zhang, Q.; Deng, M.; Miao, J.; Guo, Y.; Gao, W.; Cui, Q. An analysis of human microRNA and disease associations. *PLoS ONE* **2008**, *3*, e3420. [[CrossRef](#)] [[PubMed](#)]
37. Van Laarhoven, T.; Nabuurs, S.B.; Marchiori, E. Gaussian interaction profile kernels for predicting drug–target interaction. *Bioinformatics* **2011**, *27*, 3036–3043. [[CrossRef](#)] [[PubMed](#)]
38. Hwang, T.H.; Kuang, R. A heterogeneous label propagation algorithm for disease gene discovery. In Proceedings of the 10th SIAM International Conference on Data Mining, Columbus, OH, USA, 29 April–1 May 2010; pp. 583–594.
39. Castro, D.; Moreira, M.; Gouveia, A.M.; Pozza, D.H.; Mello, R.A.D. MicroRNAs in lung cancer. *Oncotarget* **2017**, *8*, 81679–81685. [[CrossRef](#)] [[PubMed](#)]
40. Sittka, A.; Schmeck, B. MicroRNAs in the lung. *Adv. Exp. Med. Biol.* **2013**, *774*, 121–134. [[PubMed](#)]
41. Tominaga, E.; Yuasa, K.; Shimazaki, S.; Hijikata, T. MicroRNA-1 targets Slug and endows lung cancer A549 cells with epithelial and anti-tumorigenic properties. *Exp. Cell Res.* **2013**, *319*, 77–88. [[CrossRef](#)] [[PubMed](#)]
42. Zhu, X.; Li, Y.; Shen, H.; Li, H.; Long, L.; Hui, L.; Xu, W. miR-137 inhibits the proliferation of lung cancer cells by targeting Cdc42 and Cdk6. *FEBS Lett.* **2013**, *587*, 73–81. [[CrossRef](#)] [[PubMed](#)]
43. Zheng, D.; Haddadin, S.; Wang, Y.; Gu, L.Q.; Perry, M.C.; Freter, C.E.; Wang, M.X. Plasma microRNAs as novel biomarkers for early detection of lung cancer. *Int. J. Clin. Exp. Pathol.* **2011**, *4*, 575–586. [[PubMed](#)]
44. Han, H.S.; Yun, J.; Lim, S.N.; Han, J.H.; Lee, K.H.; Kim, S.T.; Kang, M.H.; Son, S.M.; Lee, Y.M.; Choi, S.Y. Downregulation of cell-free miR-198 as a diagnostic biomarker for lung adenocarcinoma-associated malignant pleural effusion. *Int. J. Cancer* **2013**, *133*, 645–652. [[CrossRef](#)] [[PubMed](#)]
45. Hanoun, N.; Delpu, Y.; Suriawinata, A.A.; Bournet, B.; Bureau, C.; Selves, J.; Tsongalis, G.J.; Dufresne, M.; Buscail, L.; Cordelier, P. The silencing of microRNA 148a production by DNA hypermethylation is an early event in pancreatic carcinogenesis. *Clin. Chem.* **2010**, *56*, 1107–1118. [[CrossRef](#)] [[PubMed](#)]
46. Thu, K.L.; Chari, R.; Lockwood, W.W.; Lam, S.; Lam, W.L. miR-101 DNA copy loss is a prominent subtype specific event in lung cancer. *J. Thorac. Oncol.* **2011**, *6*, 1594–1598. [[CrossRef](#)] [[PubMed](#)]
47. Calin, G.A.; Sevignani, C.; Dumitru, C.D.; Hyslop, T.; Noch, E.; Yendamuri, S.; Shimizu, M.; Rattan, S.; Bullrich, F.; Negrini, M. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 2999–3004. [[CrossRef](#)] [[PubMed](#)]
48. Wang, L.K.; Hsiao, T.H.; Hong, T.M.; Chen, H.Y.; Kao, S.H.; Wang, W.L.; Yu, S.L.; Lin, C.W.; Yang, P.C. MicroRNA-133a suppresses multiple oncogenic membrane receptors and cell invasion in non-small cell lung carcinoma. *PLoS ONE* **2014**, *9*, e96765. [[CrossRef](#)] [[PubMed](#)]

49. Wei, J.; Ma, Z.; Li, Y.; Zhao, B.; Wang, D.; Jin, Y.; Jin, Y. miR-143 inhibits cell proliferation by targeting autophagy-related 2B in non-small cell lung cancer H1299 cells. *Mol. Med. Rep.* **2015**, *11*, 571–576. [[CrossRef](#)] [[PubMed](#)]
50. Ma, Z.L.; Hou, P.P.; Li, Y.L.; Wang, D.T.; Yuan, T.W.; Wei, J.L.; Zhao, B.T.; Lou, J.T.; Zhao, X.T.; Jin, Y. MicroRNA-34a inhibits the proliferation and promotes the apoptosis of non-small cell lung cancer H1299 cell line by targeting TGF β R2. *Tumour Biol.* **2015**, *36*, 2481–2490. [[CrossRef](#)] [[PubMed](#)]
51. Shang, A.Q.; Xie, Y.N.; Wang, J.; Sun, L.; Wei, J.; Lu, W.Y.; Lan, J.Y.; Wang, W.W.; Wang, L.; Wang, L.L. Predictive values of serum microRNA-22 and microRNA-126 levels for non-small cell lung cancer development and metastasis: A case-control study. *Neoplasma* **2017**, *64*, 453–459. [[CrossRef](#)] [[PubMed](#)]
52. Yang, J.S.; Li, B.J.; Lu, H.W.; Chen, Y.; Lu, C.; Zhu, R.X.; Liu, S.H.; Yi, Q.T.; Li, J.; Song, C.H. Serum miR-152, miR-148a, miR-148b, and miR-21 as novel biomarkers in non-small cell lung cancer screening. *Tumour Biol.* **2015**, *36*, 3035–3042. [[CrossRef](#)] [[PubMed](#)]
53. Joerger, M.; Baty, F.; Früh, M.; Droege, C.; Stahel, R.A.; Betticher, D.C.; Moos, R.V.; Ochsenbein, A.; Pless, M.; Gautschi, O. Circulating microRNA profiling in patients with advanced non-squamous NSCLC receiving bevacizumab/erlotinib followed by platinum-based chemotherapy at progression (SAKK 19/05). *Lung Cancer* **2014**, *85*, 306–313. [[CrossRef](#)] [[PubMed](#)]
54. Jia, Y.; Zang, A.; Shang, Y.; Yang, H.; Song, Z.; Wang, Z.; Ren, L.; Wei, Y.; Hu, L.; Shi, H. MicroRNA-146a rs2910164 polymorphism is associated with susceptibility to non-small cell lung cancer in the chinese population. *Med. Oncol.* **2014**, *31*, 1–5. [[CrossRef](#)] [[PubMed](#)]
55. Druesne-Pecollo, N.; Touvier, M.; Barrandon, E.; Chan, D.S.; Norat, T.; Zelek, L.; Hercberg, S.; Latino-Martel, P. Excess body weight and second primary cancer risk after breast cancer: A systematic review and meta-analysis of prospective studies. *Breast Cancer Res. Treat.* **2012**, *135*, 647–654. [[CrossRef](#)] [[PubMed](#)]
56. Tao, Z.; Shi, A.; Lu, C.; Song, T.; Zhang, Z.; Zhao, J. Breast cancer: Epidemiology and etiology. *Cell Biochem. Biophys.* **2015**, *72*, 333–338. [[CrossRef](#)] [[PubMed](#)]
57. Liu, J.; Mao, Q.; Liu, Y.; Hao, X.; Zhang, S.; Zhang, J. Analysis of miR-205 and miR-155 expression in the blood of breast cancer patients. *Chin. J. Cancer Res.* **2013**, *25*, 46–54. [[PubMed](#)]
58. Sun, Y.; Wang, M.; Lin, G.; Sun, S.; Li, X.; Qi, J.; Li, J. Serum microRNA-155 as a potential biomarker to track disease in breast cancer. *PLoS ONE* **2012**, *7*, e47003. [[CrossRef](#)] [[PubMed](#)]
59. Liu, R.; Li, J.; Lai, Y.; Liao, Y.; Liu, R.; Qiu, W. Hsa-miR-1 suppresses breast cancer development by down-regulating K-ras and long non-coding RNA MALAT1. *Int. J. Biol. Macromol.* **2015**, *81*, 491–497. [[CrossRef](#)] [[PubMed](#)]
60. Ouchida, M.; Kanzaki, H.; Ito, S.; Hanafusa, H.; Jitsumori, Y.; Tamaru, S.; Shimizu, K. Novel direct targets of miR-19a identified in breast cancer cells by a quantitative proteomic approach. *PLoS ONE* **2012**, *7*, e44095. [[CrossRef](#)] [[PubMed](#)]
61. Zhao, L.; Zhao, Y.; He, Y.; Mao, Y. miR-19b promotes breast cancer metastasis through targeting mylip and its related cell adhesion molecules. *Oncotarget* **2017**, *8*, 64330–64343. [[CrossRef](#)] [[PubMed](#)]
62. Cui, W.; Zhang, S.; Shan, C.; Zhou, L.; Zhou, Z. MicroRNA-133a regulates the cell cycle and proliferation of breast cancer cells by targeting epidermal growth factor receptor through the EGFR/Akt signaling pathway. *FEBS J.* **2013**, *280*, 3962–3974. [[CrossRef](#)] [[PubMed](#)]
63. Stückerath, I.; Rack, B.; Janni, W.; Jäger, B.; Pantel, K.; Schwarzenbach, H. Aberrant plasma levels of circulating miR-16, miR-107, miR-130a and miR-146a are associated with lymph node metastasis and receptor status of breast cancer patients. *Oncotarget* **2015**, *6*, 13387–13401. [[CrossRef](#)] [[PubMed](#)]
64. Lehmann, U.; Hasemeier, B.; Christgen, M.; Müller, M.; Römermann, D.; Länger, F.; Kreipe, H. Epigenetic inactivation of microRNA gene hsa-mir-9-1 in human breast cancer. *J. Pathol.* **2008**, *214*, 17–24. [[CrossRef](#)] [[PubMed](#)]

