Article

MDSCMF: Matrix Decomposition and Similarity-Constrained Matrix Factorization for miRNA–Disease Association Prediction

Jiancheng Ni [1], Lei Li [2,*], Yutian Wang [2], Cunmei Ji [2] and Chunhou Zheng [3,*]

1 Network Information Center, Qufu Normal University, Qufu 273165, China; nijch@163.com
2 School of Cyber Science and Engineering, Qufu Normal University, Qufu 273165, China; wytfuture@163.com (Y.W.); cunmeiji@126.com (C.J.)
3 School of Artifial Intelligence, Anhui University, Hefei 230601, China
* Correspondence: sdwfll1996@163.com (L.L.); zhengch99@126.com (C.Z.)

Abstract: MicroRNAs (miRNAs) are small non-coding RNAs that are related to a number of complicated biological processes, and numerous studies have demonstrated that miRNAs are closely associated with many human diseases. In this study, we present a matrix decomposition and similarity-constrained matrix factorization (MDSCMF) to predict potential miRNA–disease associations. First of all, we utilized a matrix decomposition (MD) algorithm to get rid of outliers from the miRNA–disease association matrix. Then, miRNA similarity was determined by utilizing similarity kernel fusion (SKF) to integrate miRNA function similarity and Gaussian interaction profile (GIP) kernel similarity, and disease similarity was determined by utilizing SKF to integrate disease semantic similarity and GIP kernel similarity. Furthermore, we added L_2 regularization terms and similarity constraint terms to non-negative matrix factorization to form a similarity-constrained matrix factorization (SCMF) algorithm, which was applied to make prediction. MDSCMF achieved AUC values of 0.9488, 0.9540, and 0.8672 based on fivefold cross-validation (5-CV), global leave-one-out cross-validation (global LOOCV), and local leave-one-out cross-validation (local LOOCV), respectively. Case studies on three common human diseases were also implemented to demonstrate the prediction ability of MDSCMF. All experimental results confirmed that MDSCMF was effective in predicting underlying associations between miRNAs and diseases.

Keywords: miRNA; disease; miRNA–disease association; matrix decomposition; similarity-constrained matrix factorization

1. Introduction

MiRNAs are 17–24 nt non-coding RNAs that play a pivotal role in controlling the expression of genes through RNA cleavage or translation repression [1–3]. Lin-4 was the first miRNA inspected experimentally, by Lee et al. [4] in 1993. Since that time, a large number of miRNAs have been discovered experimentally by researchers [4–6]. Researchers have found that various miRNAs are bound up with several crucial biological processes, such as cell development, cell differentiation, cell proliferation, etc. [7–10]. Developmental defects can be the result of the dysregulation of miRNAs that are associated with the progression of diseases [11]. In the meantime, numerous studies have indicated that miRNAs are connected with a serious of human neoplasms, including breast neoplasms, lung neoplasms, prostate neoplasms, etc. [12–14]. Hence, distinguishing miRNAs associated with diseases can deepen the understanding of the genetic causes of complex diseases. Strong connections between miRNAs and diseases have been found by a variety of traditional experiments in the past few years [15,16]. Traditional manual models can infer the associations between miRNAs and diseases, but these are time-consuming, laborious, and have a high failure rate. Therefore, showing the potential relationships between miRNAs and diseases requires effective and stable computational methods, which can obtain increasingly reliable miRNA–disease associations.

In the past, a great number of heterogeneous-network-based algorithms and methods have been applied to predict potential miRNA–disease relationships [17–19]. Under the assumption that miRNAs with similar functions have a high probability of being related to diseases with similar phenotypes, and vice versa [20], Jiang et al. [21] established a new calculation-based model that identified potential miRNA–disease connections by applying hypergeometric distribution. However, the similarity information utilized in this model excluded similarity scores. Li et al. [22] constructed a new model that could be used to prioritize human cancer miRNAs by measuring the associations between cancer and miRNAs based on the functional consistency scores of the miRNA target genes and the cancer-related genes. To infer the miRNA–protein connections and disease–protein connections, Mørk et al. [23] built the miPRD model. This model used selective connections to predict the relationships between miRNAs and diseases. Chen et al. [24] utilized the within and between scores of each miRNA–disease combination in the WBSMDA model to predict underlying miRNAs related to diseases. The WBSMDA model also predicted the possible relationships between new diseases and new miRNAs. Yu et al. [25] proposed an identifiable model to infer potential miRNA–disease relationships. This model combined miRNA functional similarity, disease semantic similarity and disease phenotypic similarity to create a modified information flow method. In a phenome–microRNAome network, possible connections and validated relationships between miRNAs and diseases were adopted. Chen et al. [26] introduced the Jaccard similarity between miRNAs and diseases into the BLHARMDA model to investigate prospective miRNA–disease relationships. For improving the prediction efficiency, BLHARMDA used a bipartite local model with a KNN architecture. Ha et al. [27] proposed a computational framework of metric learning named MLMD for predicting potential miRNA–disease associations. MLMD exploited distance metric learning on a miRNA–disease bipartite graph to infer unconfirmed miRNA–disease associations. The excellent performance of MLMD could be attributed to two factors: On the one hand, the implementation of metric learning overcame the violation of triangle inequality. On the other hand, the miRNA expression data were adequately trained in metric learning. Li et al. [28] proposed a similarity-constrained matrix factorization method to infer unconfirmed disease-associated miRNAs. To construct an information-rich similarity matrix, they utilized similarity network fusion to integrate various kinds of similarities. Then, similarity-based regularization terms were added to common non-negative matrix factorization to form a similarity-constrained matrix factorization algorithm, which was applied to make accuracy predictions. The above methods are mainly based on the construction of heterogeneous networks to identify and speculate on the potential disease-related miRNAs, and after cross-validation and case analysis experiments, it was proven that they can be used to observe the potential association between miRNA and disease, but their prediction performance still needs to be improved.

Recently, methods based on the random walk method have gradually been proposed, and more accurate prediction results have been obtained. Shi et al. [29] utilized the function links between human disease genes and miRNA targets to devise a novel model. A random walk algorithm and global network distance measurement were applied to search for feasible miRNA–disease relationships. Chen et al. [30] utilized a random walk with restart algorithm to construct the RWRMDA model. Because the prediction performance calculated by global network similarity was better than the of the local network [31], RWR-MDA employed global network similarity to determine the feasible interactions between microRNAs and diseases. Unfortunately, RWRMDA was inappropriate for the diseases without known associated miRNAs. Liu et al. [32] also implemented a random walk with restart algorithm in their model to make prediction results to a higher degree. They employed the random walk with restart algorithm on a heterogeneous graph established by utilizing disease similarity and miRNA similarity. Luo et al. [33] employed an imbalanced bi-random walk method on a heterogeneous network with information on miRNAs and diseases to identify feasible miRNA–disease interactions. When the random walk algorithm is used for association prediction, the initial state of disease nodes and miRNA nodes in

the network is very important. Researchers have proposed many design methods for the initial state of nodes in recent years, but the prediction performance has not been greatly improved.

As artificial intelligence technology has developed, machine-learning-based models have increasingly been employed for the accurate prediction of miRNA–disease relationships. To obtain accurate results in matrix completion for miRNA–disease association prediction, Li et al. [34] avoided using negative samples in MCMDA. To infer unknown miRNA–disease interactions, the probabilistic matrix factorization (PMF) algorithm was applied [35] to make predictions. The PMF algorithm is a machine learning technique commonly employed in recommender systems, and can effectively utilize all available data to recommend miRNAs linked to the disease in question. Ha et al. [36] utilized a matrix completion with network regularization method to recognize potential disease-related miRNAs. They considered an miRNA network as additional implicit feedback, and made predictions for disease associations with a given miRNA relying on its direct neighbors. Guo et al. [37] introduced MLPMDA—a novel model for predicting miRNA–disease associations using multilayer linear projection. They processed miRNA–disease interaction information by processing the top nearest neighbors of entities, and then used the updated miRNA–disease interactions and disease similarity to construct a heterogeneous matrix. In this heterogeneous matrix, the multilayer projection and layer-stacking strategy were employed to make predictions. However, in order to obtain dependable and steady performance, MLPMDA requires high-quality biological data. Ding et al. [38] presented a novel computational model named VGAMF for predicting miRNA–disease associations. VGAMF first integrated several different types of information about miRNAs and diseases into comprehensive similarity networks of miRNAs and diseases, which were used to extract the nonlinear representations of miRNAs and diseases based on the variational graph autoencoders. Then, VGAMF obtained the linear representations of miRNAs and diseases by implementing non-negative matrix factorization to process the miRNA–disease association matrix. Finally, a fully connected neural network combined linear representations with nonlinear representations to generate the predicted miRNA–disease association scores. Wang et al. [39] presented a novel method called NMCMDA to observe unknown disease-related miRNAs. The encoder and decoder were the two essential components in NMCMDA. The encoder was developed using a graph neural network to extract latent miRNA and disease characteristics from a heterogeneous miRNA–disease network. These latent features were used by the decoder to generate miRNA–disease association scores. For NMCMDA, a variety of encoders and decoders have been proposed. Finally, in NMCMDA, the combination of a relational graph convolutional network encoder and a neural multirelational decoder achieved the best prediction results. In summary, machine-learning-based models can produce more accurate prediction results, but most of them have difficulties in adjusting the optimal parameters and selecting negative samples, which seriously affect the training efficiency of the model.

Despite their outstandingly good performance, the abovementioned prediction models have several limitations, such as inadequate measurement of similarity, excessive noise in experimental data, and inaccurate prediction results. To overcome these limitations, we present a novel model called MDSCMF, which combines matrix decomposition with similarity-constrained matrix factorization to predict unobserved miRNA–disease associations. To construct information-rich miRNA similarity and disease similarity, we applied SKF to integrate various kinds of miRNA similarity data and disease similarity data. In addition, because the unknown miRNA–disease associations were much more numerous than the known associations, an MD algorithm was used to get rid of outliers from the miRNA–disease association matrix. Furthermore, we added L_2 regularization terms and similarity constraint terms to non-negative matrix factorization to form an SCMF algorithm, which was implemented to obtain the final association scores of each miRNA–disease pair. To evaluate the effectiveness of MDSCMF, 5-CV, global LOOCV, and local LOOCV were carried out on the known miRNA–disease association data downloaded from HMDD

v3.2 [40]. Furthermore, we performed case studies on colon neoplasms, breast neoplasms, and lung neoplasms for prediction. As a result, 29, 29, and 28 out of the top 30 miRNAs potentially connected to these high-risk human diseases, respectively, were confirmed by miR2Disease [41] and dbDEMC v2.0 [42]. Experimental results showed that MDSCMF was effective for inferring possible relationships between miRNAs and diseases.

2. Results

2.1. Performance Evaluation

In this section, based on the verified associations between miRNAs and diseases in the HMDD v3.2 database, 5-CV, global LOOCV, and local LOOCV were implemented to evaluate the prediction performance of MDSCMF.

In the framework of 5-CV, we compared MDSCMF with other previous computational methods, including GCAEMDA [43], MSCHLMDA [44], NIMCGCN [45], and HFHLMDA [46]. The full set of verified miRNA–disease associations were divided into five parts in a random manner, where the test set was held by each part in turn, while the training set consisted of the other four parts. The full set of unknown miRNA–disease associations were considered as candidate samples. We applied our method to determine the ranking of the test set relative to candidate samples. Furthermore, for the purpose of reducing potential deviations resulting in random sample segmentations, we applied 100 repeated segmentations to verify the miRNA–disease associations. When the ranking of all test samples was higher than a certain threshold, MDSCMF was regarded as a valid method. Then we could utilize the receiver operating characteristic (ROC) curve that was obtained by plotting the true positive rate (TPR) against the false positive rate (FPR) to effectively evaluate the performance of MDSCMF. We could calculate the areas under the ROC curve (AUCs) of these methods, whose values were between 0 and 1. Figure 1 indicates that MDSCMF, GCAEMDA, MSCHLMDA, NIMCGCN, and HFHLMDA had AUC values of 0.9488, 0.9415, 0.9324, 0.9378, and 0.9301, respectively. The AUC value of MDSCMF was clearly higher than that of the other methods.
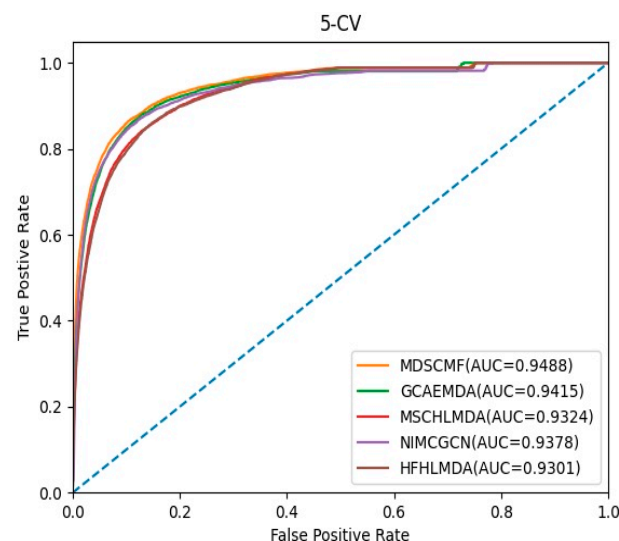

Figure 1. AUC of 5-CV compared with those of GCAEMDA, MSCHLMDA, NIMCGCN, and HFHLMDA.

In the framework of global LOOCV, MDSCMF was also compared with GCAEMDA, MSCHLMDA, NIMCGCN, and HFHLMDA. The test set was held by each verified miRNA–disease association in turn, while the training set was composed of the other verified associations. The full set of unknown miRNA–disease associations were considered as candidate samples. In addition, we applied MDSCMF to obtain all predicted association scores so that the ranking of the test set relative to the candidate samples could be determined.

Similar to 5-CV, we also calculated the AUCs of these methods so as to effectively evaluate their performance. From Figure 2, we can see that MDSCMF, GCAEMDA, MSCHLMDA, NIMCGCN, and HFHLMDA had AUC values of 0.9540, 0.9505, 0.9378, 0.9410, and 0.9321, respectively. Hence, the AUC value of MDSCMF was also higher than that of the other methods.
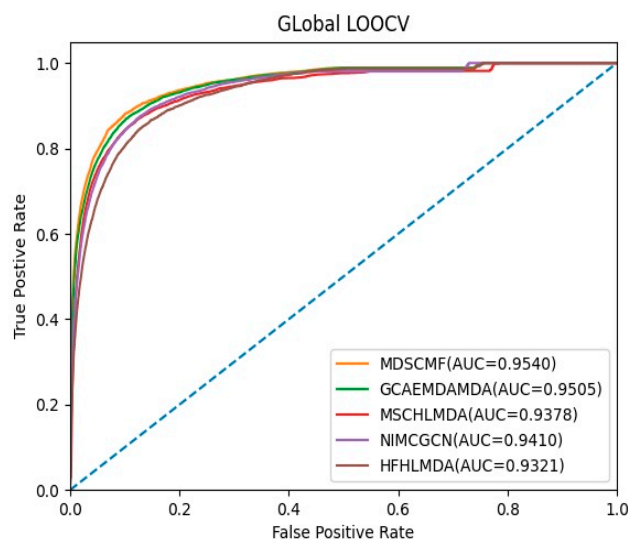

Figure 2. AUC of global LOOCV compared with those of GCAEMDA, MSCHLMDA, NIMCGCN, and HFHLMDA.

In the framework of local LOOCV, we also compared MDSCMF with other previous models (i.e., RFMDA [47], BNPMDA [48], ABMDA [49] and VGAMF [38]) to objectively evaluate its performance. In this way, we could determine the ability of MDSCMF to predict the associations between miRNAs and diseases without any verified related miRNAs. For random diseases in the HMDD v3.2 database, the confirmed associations between each disease and all miRNAs were considered as the test set, and remaining associations were regarded as the training set. Similar to the previous two cross-validation methods, the AUC value in local LOOCV still served as the evaluation criterion to reflect the ability of these models. The specific results are shown in Figure 3, which shows that the prediction performance of MDSCMF was better than that of the other models.

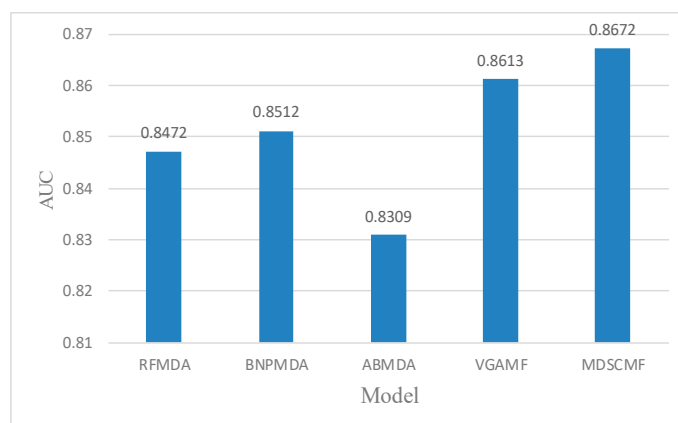**Figure 3.** Comparisons between MDSCMF and other computational models by local LOOCV.

2.2. Parameter Analysis

In this section, the parameters ϑ and σ were quantitatively analyzed to research their effects on the prediction performance. ϑ and σ were set as the regularization parameters, which were applied to control the overfitting degree and the smoothness of similarity

consistency, respectively. We utilized all combinations of two values $\vartheta \in \left\{ 2^{-3}, 2^{-2}, \ldots, 2^3 \right\}$ and $\sigma \in \left\{ 2^{-3}, 2^{-2}, \ldots, 2^3 \right\}$ to conduct MDSCMF. The AUC values of 5-CV were applied to evaluate the performance of the model under different combinations of parameters. After various tests were conducted, we concluded that the model obtained the best performance when $\vartheta = 2^2$ and $\sigma = 2^0$, as shown in Figure 4.
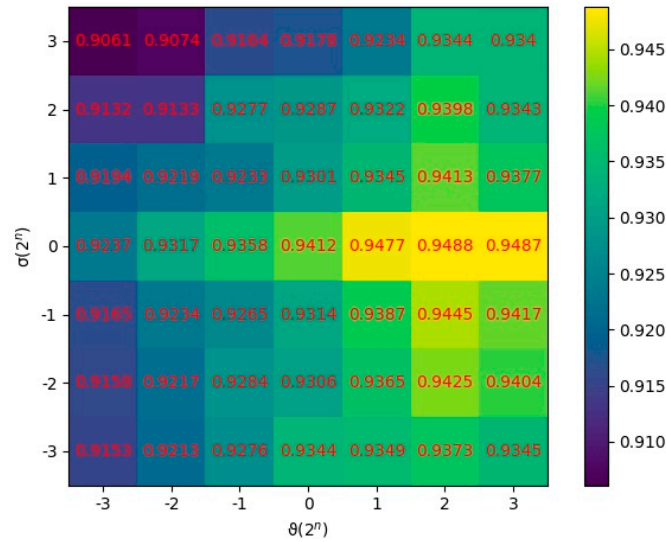

Figure 4. AUCs at different values of ϑ and σ.

2.3. Effects of Matrix Decomposition Analysis

In this section, we evaluated the effect of the pre-processing MD step for known miRNA–disease association matrix A on the model's performance. The AUC values of 5-CV were considered as indicators, and the corresponding ROC curves are shown in Figure 5. In MDSCMF, MD considers the sparsity of the miRNA–disease association matrix, thereby improving the prediction ability of the model. Conversely, MDSCMF without MD disregards the sparsity of the original association matrix; thus, the noise data in the matrix may reduce the accuracy of the prediction. As shown in Figure 5, the AUC value of MDSCMF under the 5-CV framework was 0.9488. In contrast, the AUC value of MDSCMF without MD under the 5-CV framework was 0.9291. The results of the comparison distinctly show that MDSCMF with MD has a higher AUC value compared to that without MD.
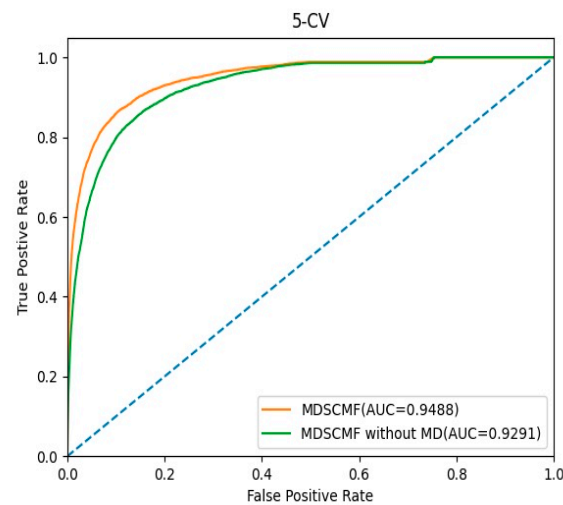

Figure 5. The ROC curves of MDSCMF and MDSCMF without MD.

2.4. Case Studies

For the purpose of demonstrating the effectiveness and accuracy of MDSCMF, we applied an evaluation experiment in this study. We implemented several types of human diseases—i.e., colon neoplasms, breast neoplasms, and lung neoplasms—as case studies to validate the performance of our method. Colon neoplasms are malignancies in the field of medicine that have been confirmed to be associated with several miRNAs [50,51]. Breast neoplasms, which have been observed to be associated with several miRNAs in clinical experiments, have a high incidence rate among women [52]. Lung neoplasms are among the most dangerous malignancies, with the fastest increases in morbidity and mortality [53]. A growing body of evidence indicates that these diseases have close relationships with several miRNAs. The miRNAs associated with these diseases were ranked in line with the prediction scores. Moreover, we utilized two databases—miR2Disease [41] and dbDEMC v2.0 [42]—to check miRNAs that had been ranked.

As a result, 29, 29, and 28 of the top 30 miRNAs inferred by our model were individually confirmed to be associated with colon neoplasms, breast neoplasms, and lung neoplasms, respectively, according to the miR2Disease [41] and dbDEMC v2.0 [42] databases. Tables 1–3 show the corresponding prediction results.

Table 1. The top 30 potential miRNAs associated with colon neoplasms.

miRNA	Evidence	miRNA	Evidence
hsa-mir-630	d	hsa-mir-29b	m; d
hsa-mir-20a	m; d	hsa-mir-141	m; d
hsa-mir-143	m; d	hsa-mir-132	m; d
hsa-mir-584	d	hsa-mir-19b	m; d
hsa-mir-506	d	hsa-mir-29a	m; d
hsa-mir-552	d	hsa-mir-223	d
hsa-mir-128	unconfirmed	hsa-let-125b	d
hsa-mir-7i	m; d	hsa-mir-622	d
hsa-mir-127	m; d	hsa-mir-18a	d
hsa-mir-1290	d	hsa-mir-143	d
hsa-mir-493	d	hsa-mir-125a	m; d
hsa-mir-498	d	hsa-mir-21	m; d
hsa-mir-107	m; d	hsa-mir-137	m; d
hsa-mir-191	m; d	hsa-mir-424	d
hsa-mir-32	m; d	hsa-mir-200b	d

m: miR2Disease database; d: dbDEMC v2.0 database.

Table 2. The top 30 potential miRNAs associated with breast neoplasms.

miRNA	Evidence	miRNA	Evidence
hsa-mir-99a	m; d	hsa-mir-663	m
hsa-mir-542	d	hsa-mir-520h	d
hsa-mir-96	d	hsa-mir-519d	d
hsa-mir-98	m; d	hsa-mir-186	d
hsa-mir-185	d	hsa-mir-381	d
hsa-mir-130a	d	hsa-mir-32	d
hsa-mir-708	d	hsa-mir-590	unconfirmed
hsa-mir-150	d	hsa-mir-330	d
hsa-mir-192	d	hsa-mir-433	d
hsa-mir-196b	d	hsa-mir-942	d
hsa-mir-888	d	hsa-mir-661	m; d
hsa-mir-9	m; d	hsa-mir-337	d
hsa-mir-130b	d	hsa-mir-494	d
hsa-mir-592	d	hsa-mir-212	d
hsa-mir-99b	d	hsa-mir-618	d

m: miR2Disease database; d: dbDEMC v2.0 database.

Table 3. The top 30 potential miRNAs associated with lung neoplasms.

miRNA	Evidence	miRNA	Evidence
hsa-mir-96	d	hsa-mir-937	unconfirmed
hsa-mir-145	m; d	hsa-mir-30e	m
hsa-mir-99a	m; d	hsa-mir-151	d
hsa-mir-9	m; d	hsa-mir-614	d
hsa-mir-185	d	hsa-mir-1323	d
hsa-mir-130a	d	hsa-mir-32	d
hsa-mir-7	m; d	hsa-mir-1298	d
hsa-mir-150	m; d	hsa-mir-330	d
hsa-mir-192	m; d	hsa-mir-433	d
hsa-mir-769	unconfirmed	hsa-mir-522	d
hsa-mir-939	d	hsa-mir-449a	d
hsa-mir-98	m; d	hsa-mir-143	m; d
hsa-mir-130b	m; d	hsa-mir-564	d
hsa-mir-638	d	hsa-mir-212	m; d
hsa-mir-99b	d	hsa-mir-615	unconfirmed

m: miR2Disease; d: dbDEMC v2.0 database.

3. Materials and Methods

In this paper, we utilized the biological information of miRNAs and diseases to propose a novel method called MDSCMF, which fully extends the advantages of matrix decomposition and similarity-constrained matrix factorization to predict possible miRNA–disease associations. The flowchart of MDSCMF is clearly shown in Figure 6.
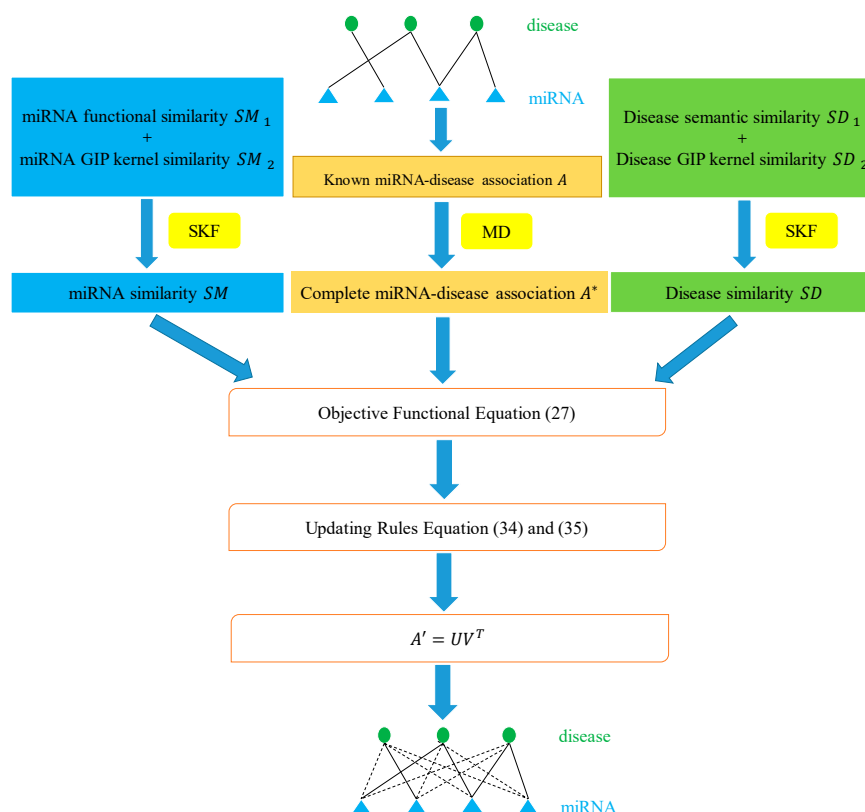

Figure 6. Flowchart of MDSCMF.

3.1. Human miRNA–Disease Associations

In this study, we took advantage of miRNA–disease association data from the HMDD v3.2 database [40], which contained 12,446 verified associations between 853 miRNAs and 591 diseases. To make calculation more convenient, we constructed an adjacency matrix

$A \in R^{nm \times nd}$ to indicate the miRNA–disease relationships. We set nd and nm to stand for the numbers of diseases and miRNAs, respectively. Specifically, the element $A(i, j)$ is equal to 1 when miRNA m_i is proved to be connected with disease d_j, and otherwise it is equal to 0. Therefore, the matrix A contains 12,446 entries that are equal to 1.

3.2. MiRNA Functional Similarity

The miRNAs with similar functions have a high probability of being related to diseases that are similar, and vice versa [20]. Therefore, we downloaded the miRNA functional similarity data from http://www.cuilab.cn/files/images/cuilab/misim.zip, accessed on 1 June 2022. For ease of calculation, we constructed the matrix SM_1 to store the data. The element $SM_1(m_i, m_j)$ represents the value of similarity between miRNA m_i and miRNA m_j.

3.3. Disease Semantic Similarity

The directed acyclic graph (DAG) based on the MeSH descriptor [54] can be utilized to describe diseases. $DAG(D) = (D, T(D), E(D))$ represents the DAG of disease D. $T(D)$ denotes the nodes in the DAG that include D itself and its ancestor nodes. $E(D)$ denotes the edges in the DAG that connect child nodes with parent nodes directly. The formula to calculate the semantic score of disease D is defined as follows:

$$DV1(D) = \sum_{d \epsilon T(D)} D_D1(d), \tag{1}$$

where the formula to calculate the contribution value $D_D1(d)$ of disease d is as follows:

$$D_D1(d) = \begin{cases} 1 & if\ d = D \\ \max\{\Delta * D_D1(d') | d' \epsilon children\ of\ d\} & if\ d \neq D' \end{cases} \tag{2}$$

where Δ is the semantic contribution factor, which was equal to 0.5 in our paper, based on previous literature [55].

The formula to obtain the semantic similarity score between disease d_i and disease d_j is defined as follows:

$$SS_1(d_i, d_j) = \frac{\sum_{t \in T(d_i) \cap T(d_j)} \left(D_{d_i}1(t) + D_{d_j}1(t) \right)}{DV1(d_i) + DV1(d_j)}. \tag{3}$$

Furthermore, for the two diseases of the same layer in a DAG, assuming they have different occurrences in DAGs, it does not make sense to define the semantic contributions of the two diseases for this DAG to be consistent. Objectively speaking, the semantic contribution of high-incidence diseases should be less than that of low-incidence diseases. Consequently, to further optimize the similarity information between diseases, another strategy was introduced to calculate disease semantic similarity following this method [56]. Specifically, the formulae to calculate the semantic score of disease D and the contribution values of disease d are as follows:

$$DV2(D) = \sum_{d \epsilon T(D)} D_D2(d), \tag{4}$$

$$D_D2(d) = -\log\left(\frac{the\ number\ of\ DAGs\ including\ d}{the\ number\ of\ diseases} \right). \tag{5}$$

Then, the formula to obtain the semantic similarity score between d_i and disease d_j is as follows:

$$SS_2(d_i, d_j) = \frac{\sum_{t \in T(d_i) \cap T(d_j)} \left(D_{d_i}2(t) + D_{d_j}2(t) \right)}{DV2(d_i) + DV2(d_j)}. \tag{6}$$

For the purpose of making the results more accurate, we set two kinds of semantic similarity that were equally important. Therefore, if disease d_i and d_j had semantic simi-

larity, we calculated the average $SD_1(d_i, d_j)$ of $SS_1(d_i, d_j)$ and $SS_2(d_i, d_j)$ by the following formula:

$$SD_1(d_i, d_j) = \frac{SS_1(d_i, d_j) + SS_2(d_i, d_j)}{2}. \tag{7}$$

3.4. Gaussian Interaction Profile Kernel Similarity

The miRNAs with similar functions have a high probability of being related to similar diseases, and vice versa [20]. Therefore, the Gaussian interaction profile kernel similarity was applied to determine the miRNA similarity and disease similarity [57,58]. We made vector $K(d_i)$ to represent the interaction profile of disease d_i in accordance with whether or not d_i had a verified association with each miRNA. Similarly, we made vector $K(m_i)$ to represent the interaction profile m_i in accordance with whether or not m_i had a verified association with each disease. The equation to calculate the GIP kernel similarity of diseases is defined as follows:

$$SD_2(d_i, d_j) = \exp\left(-\rho_d \parallel K(d_i) - K(d_j) \parallel^2\right), \tag{8}$$

where ρ_d is applied to control the kernel bandwidth. The ρ_d is obtained by normalizing the original bandwidth ρ_d' by the average number of verified associations with miRNAs per disease, as follows:

$$\rho_d = \rho_d' / (\frac{1}{nd} \sum_{i=1}^{nd} \parallel K(d_i) \parallel^2). \tag{9}$$

Similarly, we used the following equations to calculate the GIP kernel similarity of miRNAs:

$$SM_2(m_i, m_j) = \exp\left(-\rho_m \parallel K(m_i) - K(m_j) \parallel^2\right), \tag{10}$$

$$\rho_m = \rho_m' / (\frac{1}{nm} \sum_{i=1}^{nm} \parallel K(m_i) \parallel^2). \tag{11}$$

3.5. Integrating Similarity for miRNAs and Diseases

In this section, the similarity kernel fusion [59] was implemented to integrate miRNA functional similarity and GIP kernel similarity into ultimate miRNA similarity. The concrete integration process of miRNA similarity matrices can be divided into the following major steps:

In the first step, two different miRNA similarities are treated as original miRNA similarity kernels, which are defined as SM_n, $n = 1, 2$ in the above sections. Each miRNA similarity is normalized by the following equation:

$$F_n(m_i, m_j) = \frac{SM_n(m_i, m_j)}{\sum_{m_k \in M} SM_n(m_k, m_j)}, \tag{12}$$

where $F_n(m_i, m_j)$ denotes the normalized kernel that satisfies $\sum_{m_k \in M} F_n(m_i, m_j) = 1$, and $M = \{m_i\}_{i=1}^{nm}$ indicates the set of miRNAs.

In the second step, the neighbor-constraint kernel for each miRNA original kernel can be constructed as follows:

$$S_n(m_i, m_j) = \begin{cases} \frac{SM_n(m_i, m_j)}{\sum_{m_k \in N_i} SM_n(m_i, m_k)} & if \ m_j \in N_i \\ 0 & if \ m_j \notin N_i \end{cases}, \tag{13}$$

where $S_n(m_i, m_j)$ denotes a neighbor-constraint kernel that obeys $\sum_{m_k \in M} S_n(m_i, m_j) = 1$, and N_i denotes the collection of all neighbors of miRNA m_i, including itself.

In the third step, the normalized kernels and neighbor-constraint kernels are integrated as follows:

$$F_n^{l+1} = \tau \left(S_n \times \frac{\sum_{t \neq n} F_t^l}{2} \times S_n^T \right) + (1 - \tau) \frac{\sum_{t \neq n} F_t^0}{2}, \tag{14}$$

where F_n^{l+1} represents the value of n-th kernel after $l + 1$ iterations, P_t^0 represents the initial value of F_t, and the weight parameter $\tau \in (0,1)$ is used to balance the rate. After F_n^{l+1}, $n = 1, 2$ is obtained, the overall kernel SM^* can be calculated by the following formula:

$$SM^* = \frac{1}{2} \sum_{n=1}^{2} F_n^{l+1}. \tag{15}$$

In the fourth step, a weighted matrix W is applied to further eliminate noises in the overall kernel SM^*. The construction process of W is as follows:

$$W(m_i, m_j) = \begin{cases} 1 & if \ m_i \in N_j \cap m_j \in N_i \\ 0 & if \ m_i \notin N_j \cap m_j \notin N_i \\ 0.5 & otherwise \end{cases}. \tag{16}$$

In the last step, the ultimate miRNA similarity kernel $SM \in R^{nm \times nm}$ can be calculated by the following formula:

$$SM = W \times SM^*. \tag{17}$$

In the same light, we could obtain the ultimate disease similarity kernel as $SD \in R^{nd \times nd}$.

3.6. Matrix Decomposition

From the published literature [60], we found that the data used in experiments were far from perfect. Several real data of miRNA–disease associations were redundant and/or missing. Therefore, we decomposed the adjacency matrix A into two sections: The linear combination of the adjacency matrix A and low-rank matrix Y was the first section. The second section was the sparse matrix X, which included a large number of zero values. Clearly, the data of the sparse matrix X can be regarded as outliers. The matrix decomposition method was applied to acquire the lowest-rank matrix, which was employed to reconstruct a novel adjacency matrix. The formula to decompose the adjacency matrix A is defined as follows:

$$A = AY + X. \tag{18}$$

For the purpose of making the Y become low-rank, we could enforce nuclear norm on Y. In addition, the $L_{2,1}$ norm was enforced on the X so that X became sparse. The specific process can be represented by the following formula:

$$\min_{Y,X} \| Y \|_* + \varphi \| X \|_{2,1} \ s.t. \ A = AY + X, \tag{19}$$

where $\| Y \|_* = \sum_i \beta_i$ (*i.e., β_i is the sigular values of Y*) represents the nuclear norm of Y, $\| X \|_{2,1} = \sum_{j=1}^{n} \sqrt{\sum_{i=1}^{n} (X_{ij})^2}$ represents the $L_{2,1}$ norm of X, and the positive free parameter φ is applied to balance the weights of $\| Y \|_*$ and $\| X \|_{2,1}$. Furthermore, minimizing the nuclear norm of Y and the $L_{2,1}$ norm of X contributes to convenient calculation.

If the matrix A combined with Y is treated as an identity matrix, the algorithm will become the robust PCA. Therefore, Equation (19) can be treated as a robust PCA generalization [61] and changed into a comparable problem, as follows:

$$\min_{Y,X,J} \| J \|_* + \varphi \| X \|_{2,1} s.t. A = AY + X, \ Y = J. \tag{20}$$

Equation (20) is a constraint and convex optimization problem. Therefore, both the first-order information and the special properties of these convex optimization problems can be employed to solve the issue of scalability. The inexact augmented Lagrange multipliers (IALM) algorithm [62] can be utilized to convert Equation (20) to an unconstraint problem. Then, the augmented Lagrange function is adopted to minimize this problem, as follows:

$$L = \| J \|_* + \varphi \| X \|_{2,1} + tr\left(F_1^T (A - AY - X)\right) + tr\left(F_2^T (Y - J)\right) + \frac{\alpha}{2}\left(\| A - AY - X \|_F^2 + \| Y - J \|_F^2\right), \quad (21)$$

where the penalty parameter $\alpha \geq 0$. After minimization with respect to J, Y, and X, the above problem can be settled effectively. In addition, Equation (21) can be solved when the other variables are fixed and the Lagrange multipliers F_1 and F_2 are updated. The specific steps for solving Equation (21) are displayed in Algorithm 1.

Algorithm 1: Solving Equation (21) by IALM

Input: Given an incomplete matrix A and parameters $\varphi \in \{0, 1\}$
Output: Y^* and X^*
Initialize: $Y = 0$, $X = 0$, $F_1 = 0$, $F_2 = 0$, $\alpha = 10^{-4}$, $max_\alpha = 10^{10}$, $\gamma = 1.1$, $\varepsilon = 10^{-8}$
while not converged do
1: Fix the other and update J by $J = arg\,min\frac{1}{\alpha}\| J \|_* + \| \frac{1}{2}J - \left(Y + \frac{F_2}{\alpha}\right) \|_F^2$
2: Fix the other and update Y by $Y = \left(I + A^T A\right)^{-1}(A^T A - A^T X + J + \left(A^T F_1 - F_2\right)/\alpha)$
3: Fix the other and update X by $X = \arg min \frac{\varphi}{\alpha}\|X\|_{2,1} + \| \frac{1}{2}X - (AY + F_1/\alpha) \|_F^2$
4: Update the multiplier $F_1 = F_1 + \alpha(A - AY - X)$; $F_2 = F_2 + \alpha(Y - J)$
5: Update parameter α by $\alpha = \min(\gamma\alpha, max_\alpha)$
6: Check the convergence condition $||A - AY - X||_\infty < \varepsilon$ *and* $||Y - J||_\infty < \varepsilon$
end while

We defined the solution of Equation (21) as Y^* and X^*. The $A(i, j)$ was used to represent the association between miRNA m_i and disease d_j, so $Y^* \in R^{nd \times nd}$ could be applied to represent the similarity between diseases. When Y^* was obtained, the adjacency matrix A^* denoted new associations between miRNAs and diseases that could be calculated by the following equation:

$$A^* = AY^*. \quad (22)$$

3.7. Similarity-Constrained Matrix Factorization

In this section, the L_2 regularization terms and similarity constraint terms were added to a traditional non-negative matrix factorization algorithm to form similarity-constrained matrix factorization, which was applied to observe more unknown miRNA–disease interactions. The matrix $A^* \in R^{nm \times nd}$ can be factorized into $U \in R^{nm \times \gamma}$ and $V \in R^{nd \times \gamma}$, where γ represents the dimensions of miRNA features and disease features. Concretely, the miRNA–disease association can be regarded as the inner product between the miRNA feature vector and the disease feature vector: $a^*_{ij} \approx u_i v_j^T$, where a^*_{ij} indicates the (i, j)th element of matrix A^*, while u_i and v_j indicate the ith row of U and the jth row of V, respectively. The corresponding objective function is defined as follows:

$$min\frac{1}{2}\sum_{ij}\left(a^*_{ij} - u_i v_j^T\right)^2. \quad (23)$$

In what follows, the L_2 regularization terms of u_i and v_j are added to above function for preventing overfitting in the model:

$$min\frac{1}{2}\sum_{ij}\left(a^*_{ij} - u_i v_j^T\right)^2 + \frac{\vartheta}{2}\sum_{i}\| u_i \|^2 + \frac{\vartheta}{2}\sum_{j}\| v_j \|^2, \quad (24)$$

where ϑ denotes the regularization parameter for controlling the balance.

When data points are mapped from high-rank space into low-rank space, the geometric properties of the data points will most likely stay the same [63,64]. Owing to the miRNA similarity SM and disease similarity SD being able to represent the geometric structure of the data points, the similarity constraint terms S_U and S_V are proposed as follows:

$$S_U = \frac{1}{2} \sum_{ij} \| u_i - u_j \|^2 SM_{ij}, \tag{25}$$

$$S_V = \frac{1}{2} \sum_{ij} \| v_i - v_j \|^2 SD_{ij}, \tag{26}$$

where SM_{ij} represents the similarity between miRNAs m_i and m_j, while SD_{ij} denotes the similarity between diseases d_i and d_j. Because the degree of similarity between two random data points is determined by the distance between them, S_U will incur a heavy penalty if the distance between m_i and m_j is close in the miRNA feature space. Thus, we minimized the S_U to keep the geometric structure of the miRNA data points, which would give rise to m_i and m_j being mapped closely in low-dimensional space. The same is true for disease data nodes, so we also minimized the S_V. Based on the above analysis, the objective function of SCMF can be defined by adding S_U and S_V to Equation (24) as follows:

$$\min_{U,V} L = \frac{1}{2} \sum_{ij} \left(a^*_{ij} - u_i v_j^T \right)^2 + \frac{\vartheta}{2} \sum_i \| u_i \|^2 + \frac{\vartheta}{2} \sum_j \| v_j \|^2 + \frac{\sigma}{2} \sum_{ij} \| u_i - u_j \|^2 SM_{ij} + \frac{\sigma}{2} \sum_{ij} \| v_i - v_j \|^2 SD_{ij}, \tag{27}$$

where σ denotes the hyperparameter to control the smoothness degree of similarity consistency. Subsequently, an efficacious optimization algorithm is proposed to calculate the objective function of SCMF.

First, the partial derivatives of L with respect to u_i and v_j can be calculated by the following formulae:

$$\begin{aligned} \nabla_{u_i} L &= \sum_j \left(u_i v_j^T - a^*_{ij} \right) v_j + \vartheta u_i + \sigma \left(\sum_j (u_i - u_j) SM_{ij} - \sum_j (u_j - u_i) SM_{ji} \right) \\ &= u_i \left(V^T V + \vartheta I + \sigma \left(\sum_j SM_{ij} + \sum_j SM_{ji} \right) I \right) - A^*(i,:)V - \sigma \sum_j (SM_{ij} + SM_{ji}) u_j, \end{aligned} \tag{28}$$

$$\begin{aligned} \nabla_{v_j} L &= \sum_j (v_j u_i^T - a^*_{ij}) u_i + \vartheta v_j + \sigma \left(\sum_i (v_j - v_i) SD_{ji} - \sum_i (v_i - v_j) SD_{ij} \right) \\ &= v_j \left(U^T U + \vartheta I + \sigma \left(\sum_i SD_{ij} + \sum_i SD_{ji} \right) I \right) - A^*(:,j)^T U - \sigma \sum_i (SD_{ij} + SD_{ji}) v_i, \end{aligned} \tag{29}$$

where $A^*(i,:)$ and $A^*(:,j)$ indicate the ith row and jth column of matrix A^*, respectively.

Next, the calculation of the second derivatives of L with respect to u_i and v_j is presented as follows:

$$\nabla^2_{u_i} L = V^T V + \vartheta I + \sigma \left(\sum_j SM_{ij} + \sum_j SM_{ji} \right) I, \tag{30}$$

$$\nabla^2_{v_j} L = U^T U + \vartheta I + \sigma \left(\sum_i SD_{ij} + \sum_i SD_{ji} \right) I. \tag{31}$$

Then, u_i and v_j can be iteratively updated according to Newton's method, as follows:

$$u_i \leftarrow u_i - \nabla_{u_i} L \left(\nabla^2_{u_i} L \right)^{-1}, \tag{32}$$

$$v_j \leftarrow v_j - \nabla_{v_j} L \left(\nabla^2_{v_j} L \right)^{-1}. \tag{33}$$

More specifically, the update of u_i and v_j can be performed by the below formulas:

$$u_i \leftarrow \left(A^*(i,:)V + \sigma \sum_j (SM_{ij} + SM_{ji})u_j \right) \left(V^T V + \vartheta I + \sigma \left(\sum_j SM_{ij} + \sum_j SM_{ji} \right) I \right)^{-1}, \tag{34}$$

$$v_j \leftarrow \left(A^*(:,j)^T U + \sigma \sum_i (SD_{ij} + SD_{ji})v_i \right) \left(U^T U + \vartheta I + \sigma \left(\sum_i SD_{ij} + \sum_i SD_{ji} \right) I \right)^{-1}. \tag{35}$$

The update of u_i and v_j will stop when the convergence condition is satisfied. After that, the predicted association matrix can be calculated by the following formula:

$$A' = UV^T. \tag{36}$$

The value of A'_{ij} denotes the predicted association score between miRNA m_i and disease d_j. The higher the prediction score, the greater the association probability.

4. Discussion

To solve the problems of inadequate measurement of similarity, excessive noise in experimental data, and inaccurate prediction results existing in previous prediction models, we developed a computational model for predicting miRNA–disease associations based on matrix decomposition and similarity-constrained matrix factorization (MDSCMF). Because the miRNA–disease association matrix was a sparse matrix, we applied the MD algorithm to complete it. Our results demonstrated that the MD algorithm could improve the prediction performance to some extent. In addition, we applied SKF to integrate various types of similarities for constructing information-rich miRNA similarity and disease similarity. Furthermore, L_2 regularization terms and similarity constraint terms were added to non-negative matrix factorization to form the SCMF algorithm, which was utilized to generate association scores of each miRNA–disease pair. In the frameworks of 5-CV, global LOOCV, and local LOOCV, the AUCs of MDSCMF achieved 0.9488, 0.9540, and 0.8672, respectively, indicating that the performance of our method had a significant improvement relative to previous methods. Furthermore, the predicted miRNAs related to colon neoplasms, prostate neoplasms, and lung neoplasms were confirmed by the experimental literature, so the prediction results generated by our method were proven to be reliable.

It should be noted that the following factors may contribute to the reliable performance of MDSCMF: First of all, the MD algorithm, which greatly alleviated the influence of the inherent noise existing in the current dataset, was utilized to refine the miRNA–disease association matrix. In addition, when we used SCMF to make predictions, the L_2 regularization terms and similarity constraint terms could avoid overfitting problems and generate robustness of the data richness, respectively.

However, several limitations may influence the performance of MDSCMF. First of all, although the amount of data had increased, we still ought to spare no effort to expand the experimental data. Furthermore, the data we utilized included miRNA function similarity data and disease semantic similarity data, which may contain noise and outliers. Therefore, we should continuously optimize our model to improve its performance in the future.

Author Contributions: Conceptualization, J.N. and L.L.; methodology, J.N.; validation, C.J., Y.W. and C.Z.; investigation, J.N.; data curation, L.L.; writing—original draft preparation, J.N.; writing—review and editing, L.L. and C.Z.; supervision, L.L.; funding acquisition, J.N., Y.W. and C.Z.; All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data are present within the manuscript or available by request to corresponding author, Lei Li (sdwfll1996@163.com).

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

References

1. Vasques, L.R.; Pereira, L.V.; Izzotti, A.; Schoof, C.R.G.; Ribeiro, A.O. MicroRNAs: Modulators of cell identity, and their applications in tissue engineering. *MicroRNA* **2014**, *3*, 45–53.
2. Bartel, D.P. MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. *Cell* **2004**, *116*, 281–297. [CrossRef]
3. He, L.; Hannon, G.J. MicroRNAs: Small RNAs with a big role in gene regulation. *Nat. Rev. Genet.* **2004**, *5*, 522–531. [CrossRef] [PubMed]
4. 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]
5. Ana, K.; Sam, G.J. miRBase: Annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* **2014**, *42*, 68–73.
6. Jopling, C.L.; Yi, M.; Lancaster, A.M.; Lemon, S.M.; Sarnow, P. Modulation of hepatitis C virus RNA abundance by a liver-specific MicroRNA. *Science* **2005**, *309*, 1577–1581. [CrossRef]
7. Cheng, A.M.; Byrom, M.W.; Shelton, J.; Ford, L.P. Antisense inhibition of human miRNAs and indications for an involvement of miRNA in cell growth and apoptosis. *Nucleic Acids Res.* **2005**, *33*, 1290–1297. [CrossRef] [PubMed]
8. Karp, X.; Ambros, V. Developmental biology. Encountering MicroRNAs in Cell Fate Signaling. *Science* **2005**, *310*, 1288–1289. [CrossRef] [PubMed]
9. Miska, E.A. How microRNAs control cell division, differentiation and death. *Curr. Opin. Genet. Dev.* **2005**, *15*, 563–568. [CrossRef]
10. Xu, P.; Guo, M.; Hay, B.A. MicroRNAs and the regulation of cell death. *Trends Genet.* **2004**, *20*, 617–624. [CrossRef]
11. Lynam-Lennon, N.; Maher, S.; Reynolds, J.V. The roles of microRNA in cancer and apoptosis. *Biol. Rev. Camb. Philos. Soc.* **2009**, *84*, 55–71. [CrossRef]
12. Meola, N.; Gennarino, V.A.; Banfi, V. microRNAs and genetic diseases. *PathoGenetics* **2009**, *2*, 7. [CrossRef]
13. Yanaihara, N.; Caplen, N.; Bowman, E.; Seike, M.; Kumamoto, K. Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell* **2006**, *9*, 189–198. [CrossRef]
14. Yanaihara, N.; Caplen, N.; Bowman, E.; Seike, M.; Kumamoto, K.; Yi, M.; Stephens, R.M.; Okamoto, A.; Yokota, J.; Tanaka, T.; et al. Circulating microRNAs as potential new biomarkers for prostate cancer. *Cancer Cell* **2013**, *108*, 1925–1930.
15. Thomson, J.M.; Parker, J.S.; Hammond, S.M. Microarray Analysis of miRNA Gene Expression. *Methods Enzymol.* **2007**, *427*, 107–122. [CrossRef]
16. Han, K.; Xuan, P.; Ding, J.; Zhao, Z.J.; Hui, L.; Zhong, Y.L. Prediction of disease-related microRNAs by incorporating functional similarity and common association information. *Genet. Mol. Res.* **2014**, *13*, 2009–2019. [CrossRef]
17. You, Z.; Huang, Z.; Zhu, Z.; Yan, G.; Li, Z.; 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] [PubMed]
18. Chen, X.; Gong, Y.; Zhang, D.-H.; You, Z.; Li, Z.-W. DRMDA: Deep representations-based miRNA-disease association prediction. *J. Cell. Mol. Med.* **2018**, *22*, 472–485. [CrossRef]
19. Zeng, X.; Xuan, X.; Quan, B. Integrative approaches for predicting microRNA function and prioritizing disease-related mi-croRNA using biological interaction networks. *Brief Bioinform.* **2016**, *17*, 193–203. [CrossRef] [PubMed]
20. Lu, M.; Zhang, Q.; Min, D.; Jing, M.; Guo, Y.; Guo, W.; Cui, Q. An Analysis of Human MicroRNA and Disease Associations. *PLoS ONE* **2008**, *3*, e3420. [CrossRef]
21. Jiang, Q.; Hao, Y.; Wang, G.; Juan, L.; Wang, Y. Prioritization of disease microRNAs through a human phenome-microRNAome network. *BMC Syst. Biol.* **2010**, *4*, S2. [CrossRef] [PubMed]
22. Li, X.; Wang, Q.; Zheng, Y.; Lv, S.; Ning, S.; Sun, J.; Huang, T.; Zheng, Q.; Ren, H.; Xu, J.; et al. Prioritizing human cancer microRNAs based on genes' functional consistency between microRNA and cancer. *Nucleic Acids Res.* **2011**, *39*, e153. [CrossRef] [PubMed]
23. Mørk, S.; Sune, P.F.; Albert, P.C.; Jan, G.; Jensen, L.J. Protein-driven inference of miRNA–disease associations. *Bioinformatics* **2014**, *30*, 392–397. [CrossRef]
24. Chen, X.; Yan, C.; Zhang, X.; You, Z.; Deng, L.; Liu, Y.; Zhang, Y.; Dai, Q. WBSMDA: Within and Between Score for MiRNA-Disease Association prediction. *Sci. Rep.* **2016**, *6*, 21106. [CrossRef] [PubMed]
25. Yu, H.; Chen, X.; Lu, L. Large-scale prediction of microRNA-disease associations by combinatorial prioritization algorithm. *Sci. Rep.* **2017**, *7*, 43792. [CrossRef] [PubMed]
26. Chen, X.; Cheng, J.; Yin, J. Predicting microRNA-disease associations using bipartite local models and hubness-aware regression. *RNA Biol.* **2018**, *15*, 1192–1205. [CrossRef] [PubMed]
27. Ha, J.; Park, C. MLMD: Metric Learning for predicting MiRNA-Disease associations. *IEEE Access.* **2021**, *9*, 78847–78858. [CrossRef]
28. Li, L.; Gao, Z.; Wang, Y.; Zhang, M.; Ni, J.; Zheng, C.; Su, Y. SCMFMDA: Predicting microRNA-disease associations based on similarity constrained matrix factorization. *PLoS Comput. Biol.* **2021**, *17*, e1009165. [CrossRef]

29. Shi, H.; Xu, J.; Zhang, G.; Xu, L.; Li, C.; Wang, L.; Zhao, Z.; Jiang, W.; Guo, Z.; Li, X. Walking the interactome to identify human miRNA-disease associations through the functional link between miRNA targets and disease genes. *BMC Syst. Biol.* **2013**, *7*, 101. [CrossRef]

30. Chen, X.; Liu, M.; Yan, G. RWRMDA: Predicting novel human microRNA-disease associations. *Mol. Biosyst.* **2012**, *8*, 2792–2798. [CrossRef]

31. Köhler, S.; Bauer, S.; Horn, D.; Robinson, P.N. Walking the Interactome for Prioritization of Candidate Disease Genes. *Am. J. Hum. Genet.* **2008**, *82*, 949–958. [CrossRef]

32. Liu, Y.; Zeng, X.; He, Z.; Zou, Q. Inferring microRNA-disease associations by random walk on a heterogeneous network with multiple data sources. *ACM Trans. Comput. Biol. Bioinform.* **2017**, *14*, 905–915. [CrossRef] [PubMed]

33. Luo, J.; Xiao, Q. A novel approach for predicting microRNA-disease associations by unbalanced bi-random walk on heterogeneous network. *J. Biomed. Inform.* **2017**, *66*, 194–203. [CrossRef]

34. Li, J.; Rong, Z.; Chen, X.; Yan, G.; You, Z. MCMDA: Matrix completion for MiRNA-disease association prediction. *Oncotarget* **2017**, *8*, 21187–21199. [CrossRef]

35. Xu, J.; Cai, L.; Liao, B.; Zhu, W.; Yang, J. Identifying Potential miRNAs-Disease Associations with Probability Matrix Factorization. *Front. Genet.* **2019**, *10*, 1234. [CrossRef] [PubMed]

36. Hua, J.; Park, C.; Park, C.; Park, S. Improved Prediction of miRNA-Disease Associations Based on Matrix Completion with Network Regularization. *Cells* **2020**, *9*, 881. [CrossRef]

37. Guo, L.; Shi, K.; Lin, W. MLPMDA: Multi-layer linear projection for predicting miRNA-disease association. *Knowl. Based Syst.* **2021**, *214*, 106718. [CrossRef]

38. Ding, Y.; Lei, X.; Liao, B.; Wu, F. Predicting miRNA-Disease Associations Based on Multi-View Variational Graph Auto-Encoder with Matrix Factorization. *IEEE J. Biomed. Health Inform.* **2021**, *26*, 446–457. [CrossRef] [PubMed]

39. Wang, J.; Li, J.; Yue, K.; Wang, L.; Ma, Y.; Li, Q. NMCMDA: Neural multicategory MiRNA-disease association prediction. *Brief Bioinform.* **2021**, *22*, bbab074. [CrossRef] [PubMed]

40. Huang, Z.; Shi, J.; Gao, Y.; Cui, C.; Zhang, S.; Li, J.; Zhou, Y.; Cui, Q. HMDD v3. 0: A database for experimentally supported human microRNA-disease associations. *Nucleic Acids Res.* **2019**, *47*, D1013–D1017. [CrossRef]

41. 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]

42. Yang, Z.; Wu, L.; Wang, A.; Tang, W.; Zhao, Y.; Zhao, H.; Teschendorff, A.E. dbDEMC 2.0: Updated database of differ-entially expressed miRNAs in human cancers. *Nucleic Acids Res.* **2017**, *45*, D812–D818. [CrossRef] [PubMed]

43. Li, L.; Wang, Y.; Ji, C.; Zheng, C.; Ni, J.; Su, Y. GCAEMDA: Predicting miRNA-disease associations via graph convolutional. *PLoS Comput. Biol.* **2021**, *17*, e1009655. [CrossRef]

44. Wu, Q.; Wang, Y.; Gao, Z.; Ni, J.; Zheng, C. MSCHLMDA: Multi-Similarity Based Combinative Hypergraph Learning for Predicting MiRNA-Disease Association. *Front. Genet.* **2020**, *11*, 354. [CrossRef]

45. Li, J.; Zhang, S.; Liu, T.; Ning, C.; Zhang, Z.; Zhou, W. Neural Inductive Matrix Completion with Graph Convolutional Networks for miRNA-disease Association Prediction. *Bioinformatics* **2020**, *36*, 2538–2546. [CrossRef]

46. Wang, Y.; Wu, Q.; Gao, Z.; Ni, J.; Zheng, C. MiRNA-disease association prediction via hypergraph learning based on high-dimensionality features. *BMC Med. Inform. Decis. Mak.* **2021**, *21*, 133. [CrossRef] [PubMed]

47. Chen, X.; Wang, C.; Yin, J.; You, Z. Novel Human miRNA-Disease Association Inference Based on Random Forest. *Mol. Ther. -Nucleic Acids* **2018**, *13*, 568–579. [CrossRef] [PubMed]

48. Chen, X.; Xie, D.; Wang, L.; Zhao, Q.; You, Z.; Liu, H. BNPMDA: Bipartite Network Projection for MiRNA–Disease Association prediction. *Bioinformatics* **2018**, *34*, 3178–3186. [CrossRef]

49. Zhao, Y.; Chen, X.; Yin, J. Adaptive boosting-based computational potential miRNA-disease associations. *Bioinformatics* **2020**, *36*, 330. [CrossRef]

50. DeSantis, C.E.; Miller, K.D.; Sauer, A.G.; Jemal, A.; Siegel, R.L. Cancer statistics for African Americans, 2019. *CA Cancer J. Clin.* **2019**, *69*, S7. [CrossRef]

51. Thackeray, E.W.; Charatcharoenwitthaya, P.; Elfaki, D.; Sinakos, E.; Lindor, K.D. Colon Neoplasms Develop Early in the Course of Inflammatory Bowel Disease and Primary Sclerosing Cholangitis. *Clin. Gastroenterol. Hepatol.* **2011**, *9*, 52–56. [CrossRef] [PubMed]

52. Fu, S.W.; Chen, L.; Man, Y.-G. miRNA Biomarkers in Breast Cancer Detection and Management. *J. Cancer* **2011**, *2*, 116–122. [CrossRef] [PubMed]

53. Wang, X.; Yin, H.; Zhang, L.; Zhang, D.; Zhu, J. The construction and analysis of the aberrant lncRNA-miRNA-mRNA network in non-small cell lung cancer. *J. Thorac. Dis.* **2019**, *11*, 1772–1778. [CrossRef]

54. Lipscomb, C.E. Medical Subject Headings (MeSH). *Bull. Med. Libr. Assoc.* **2000**, *88*, 265–266.

55. Chen, X.; Yin, J.; Qu, J.; Huang, L.; Wang, E. MDHGI: Matrix Decomposition and Heterogeneous Graph Inference for miRNA-disease association prediction. *PLoS Comput. Biol.* **2018**, *14*, e1006418. [CrossRef]

56. Xuan, P.; Han, K.; Guo, M.; Guo, Y.; Huang, Y. Prediction of microRNAs Associated with Human Diseases Based on Weighted k Most Similar Neighbors. *PLoS ONE* **2013**, *8*, e70204. [CrossRef]

57. Van, L.T.; Nabuurs, S.B.; Marchiori, E. Gaussian inter-action profile kernels for predicting drug-target interaction. *Bioinformatics* **2011**, *27*, 3036–3043.

58. Chen, X.; Huang, Y.-A.; You, Z.-H.; Yan, G.-Y.; Wang, X. A novel approach based on KATZ measure to predict associations of human microbiota with non-infectious diseases. *Bioinformatics* **2017**, *33*, 733–739. [CrossRef] [PubMed]
59. Wang, B.; Mezlini, A.M.; Demir, F.; Fiume, M.; Tu, Z.W.; Brudno, M.; Haibe-Kains, B.; Goldenberg, A. Similarity network fusion for aggregating data types on a genomic scale. *Nat. Methods* **2014**, *11*, 333–337. [CrossRef]
60. Yu, S.; Liang, C.; Xiao, Q.; Li, G.; Ding, P.; Luo, J. MCLPMDA: A novel method for miRNA-disease association prediction based on matrix completion and label propagation. *J. Cell. Mol. Med.* **2019**, *23*, 1427–1438. [CrossRef] [PubMed]
61. Chen, J.; Yang, J. Robust Subspace Segmentation Via Low-Rank Representation. *IEEE Trans. Cybern.* **2014**, *44*, 1432–1445. [CrossRef]
62. Meng, F.; Yang, X.; Zhou, C. The Augmented Lagrange Multipliers Method for Matrix Completion from Corrupted Samplings with Application to Mixed Gaussian-Impulse Noise Removal. *PLoS ONE* **2014**, *9*, e108125. [CrossRef] [PubMed]
63. Zhang, W.; Liu, X.R.; Chen, Y.L.; Wu, W.J.; Wang, W.; Li, X.H. Feature-derived graph regularized matrix factorization for predicting drug side effects-Science Direct. *Neurocomputing* **2018**, *287*, 154–162. [CrossRef]
64. Rana, B.; Juneja, A.; Saxena, M.; Gudwani, S.; Kumaran, S.S.; Behari, M.; Agrawal, R.K. Graph-Theory-based Spectral Feature Selection for Computer Aided Diagnosis of Parkinson's Disease Using T1-weighted MRI. *Int. J. Imaging Syst. Technol.* **2015**, *25*, 245–255. [CrossRef]