MCMDA: Matrix completion for MiRNA-disease association prediction

Jian-Qiang Li[1,*], Zhi-Hao Rong[2,*], Xing Chen[3], Gui-Ying Yan[4], Zhu-Hong You[5]

[1]College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, 518060, China

[2]School of Software, Beihang University, Beijing, 100191, China

[3]School of Information and Control Engineering, China University of Mining and Technology, Xuzhou, 221116, China

[4]Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, 100190, China

[5]Xinjiang Technical Institute of Physics and Chemistry, Chinese Academy of Science, ÜRÜMQI, 830011, China

[*]The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint first authors

Correspondence to: Xing Chen, *email:* xingchen@amss.ac.cn

ABSTRACT

Nowadays, researchers have realized that microRNAs (miRNAs) are playing a significant role in many important biological processes and they are closely connected with various complex human diseases. However, since there are too many possible miRNA-disease associations to analyze, it remains difficult to predict the potential miRNAs related to human diseases without a systematic and effective method. In this study, we developed a Matrix Completion for MiRNA-Disease Association prediction model (MCMDA) based on the known miRNA-disease associations in HMDD database. MCMDA model utilized the matrix completion algorithm to update the adjacency matrix of known miRNA-disease associations and furthermore predict the potential associations. To evaluate the performance of MCMDA, we performed leave-one-out cross validation (LOOCV) and 5-fold cross validation to compare MCMDA with three previous classical computational models (RLSMDA, HDMP, and WBSMDA). As a result, MCMDA achieved AUCs of 0.8749 in global LOOCV, 0.7718 in local LOOCV and average AUC of 0.8767+/-0.0011 in 5-fold cross validation. Moreover, the prediction results associated with colon neoplasms, kidney neoplasms, lymphoma and prostate neoplasms were verified. As a consequence, 84%, 86%, 78% and 90% of the top 50 potential miRNAs for these four diseases were respectively confirmed by recent experimental discoveries. Therefore, MCMDA model is superior to the previous models in that it improves the prediction performance although it only depends on the known miRNA-disease associations.

INTRODUCTION

MicroRNA (miRNA) is a kind of short non-coding single-stranded RNA (~22nt) which can regulate the gene expression by binding to the 3' untranslated regions (UTRs) of its target messenger RNA (mRNA) through base pairing [1, 2]. There are significant differences between the miRNAs in different tissues and different growth stages, which means that miRNAs have differential spatial and temporal expression patterns [3]. Based on plenty of biological experiments, researchers now believe that these small molecules have a wide range of regulation effects on eukaryotic gene expression, not only in human genes but also in genes of many other species [4]. Up to now, researchers have discovered that miRNAs are involved in a series of critical life processes, including early cell growth, proliferation, differentiation [5, 6], apoptosis, death [7], fat metabolism and so on. Therefore, it is no wonder that miRNAs are closely related to many complex human diseases [8, 9]. For example, studies have implicated that miRNA-143 and miRNA-145 are constantly down-regulated in colorectal

tumors [10] and recently Croce et al. also have shown that the downregulation of these miRNAs is a common occurrence in breast carcinomas [11]. Besides, studies by Takamizawa et al. [12] and Yanaihara et al. [13] have presented evidence that transcripts of certain let-7 homologs are significantly downregulated in human lung cancer. Based on real-time polymerase chain reaction (PCR), the analysis of miRNA arrays using pooled RNA samples from five gastric cancer patients indicates that the expression of miRNA-107, miRNA-21, miRNA-196a, miRNA-26b, miRNA-9, miRNA-142-3p, miRNA-30b, miRNA-150, miRNA-191, and miRNA-17 was found to be upregulated [14]. However, it is expensive and time-consuming to identify the associations between miRNAs and diseases using experimental methods. Considering that large numbers of miRNA-associated datasets are available, computational methods are efficient in predicting miRNA-disease associations in that they can select the most promising associated miRNAs for further experimental studies [15–17]. Therefore, it is necessary for us to make further efforts and develop efficient computational models to predict the potential miRNA-disease associations [16, 18–31].

Many computational methods have been established to predict the potential associations between miRNAs and diseases depending on the assumption that miRNAs with similar functions are more likely to have connections with diseases which share similar phenotypes [32, 33]. Jiang et al. [34] proposed a hypergeometric distribution-based model to predict miRNA-disease associations based on disease phenotype similarity network, miRNA functional similarity network, and known human disease-miRNA association network. However, this method strongly depends on the miRNA-target interactions with a high rate of false positive and false negative samples. Moreover, Shi et al. [35] presented a new model by implementing random walk algorithm on protein-protein interaction (PPI) network based on the idea that miRNAs whose target genes are related to certain diseases are more likely to be associated with these diseases. They made use of the miRNA–target interactions, disease–gene associations, and PPIs to acquire potential associations between the miRNAs and diseases. Mork et al. [36] proposed a miRPD method with the help of protein-disease interactions as well as protein-miRNA interactions, where not only disease-related miRNAs but also potential disease-related proteins were analyzed. By integrating known disease–gene associations and miRNA-target interactions, Xu et al. [37] introduced a miRNA prioritization method which need not rely on the known miRNA-disease associations. Instead, what they needed to do was to evaluate the similarity between the targets of miRNAs and disease genes. Nevertheless, all the methods mentioned above suffered from the miRNA-target interactions with high false positive and false negative samples, which could

significantly reduce the accuracy of the aforementioned models.

Researchers also proposed some other computational models without relying on miRNA-target interactions. Based on miRNA functional similarity, disease semantic similarity, disease phenotype similarity, and miRNA-disease associations, Xuan et al. [38] presented an HDMP model which analyzed the miRNAs related to the diseases by considering the functional similarities of the miRNA's k most similar neighbors. Compared with the previous methods, HDMP assigned higher weight to the miRNAs in the cluster and family since they are more likely to be associated with similar diseases. When applied to new diseases without some known related miRNAs, however, HDMP is unable to work since it strongly depends on the neighbors of the miRNAs. Besides, HDMP is based on a local similarity measure rather than a global measure which can notably promote the prediction performance. Xuan et al. [39] introduced another model called MIDP based on random walk, which exploited the characteristics of the nodes and the various ranges of topologies. The labeled nodes in MIDP were assigned higher transition weight than the unlabeled nodes, which efficiently exploited the prior information of nodes and various ranges of topologies. What is worth mentioning is that MIDP effectively relieved the negative effect of noisy data. MIDP also extended the walk on a miRNA-disease bilayer network to predict candidate specially for the diseases without any known miRNAs. Recently, Zeng et al. [40] utilized matrix completion to predict the miRNA-disease associations based on miRNA-miRNA network and disease-disease network. The method contributed multiple feature sets to address problems related to insufficient miRNA-disease association data. The method could be applied to predict unknown miRNA-disease associations and new pathogenic miRNAs for well-characterized diseases. Chen et al. [41] proposed RWRMDA model which integrated miRNA-miRNA functional similarity and known miRNA-disease associations information to predict miRNA-disease associations. RWRMDA was motivated based on the investigation that global similarity measures are better in predicting the associations between miRNAs and diseases than the previous local network similarity measures. Still, this method fails to predict miRNAs associated with new diseases without any known related miRNAs. Chen et al. [16] presented another model called WBSMDA based on miRNA functional similarity, disease semantic similarity, miRNA-disease associations, and Gaussian interaction profile kernel similarity for miRNAs and diseases. WBSMDA makes a breakthrough in that it succeeds in predicting related miRNAs for new diseases without known related miRNAs and new miRNAs without known related diseases. Recently, Chen et al. [42] presented a model of HGIMDA using miRNA functional similarity, disease semantic similarity,

miRNA-disease associations, and Gaussian interaction profile kernel similarities. In HGIMDA, the new miRNA functional similarity network was obtained by combining miRNA functional similarity network with Gaussian interaction profile kernel similarities for miRNAs. The process of calculating new disease similarity network was quite similar. Then, a heterogeneous graph was obtained by combining new miRNA functional similarity network, new disease similarity network and known miRNA-disease associations. Moreover, the potential association between a disease and a miRNA could be inferred based on an iterative equation if they didn't have known association. It has been verified that HGIMDA obtained a high prediction performance.

In addition, several computational models have considered machine learning methods. For instance, Xu et al. [43] developed a miRNA target-dysregulated network (MTDN) based on miRNA-target interactions as well as miRNA and mRNA expression profiles. Besides, MTDN implemented support vector machine (SVM) classifier to distinguish positive miRNA-disease associations from negative ones. Nevertheless, it is still fairly difficult to obtain the negative miRNA-disease associations today, which seriously decreases the prediction performance of this computational model. Chen et al. [15] presented a RLSMDA model based on semi-supervised learning which calculated the semantic similarity between different diseases. It is worth mentioning that RLSMDA could identify related miRNAs for diseases without any known associated miRNAs, meanwhile avoiding the problem of using negative associations between miRNAs and diseases. The trouble of RLSMDA is how to find the appropriate parameters and how to combine the classifiers from miRNA space and disease space together. Chen et al. [19] developed another computational model called RBMMMDA based on miRNA-disease associations which presented restricted Boltzmann machine (RBM) which is a two-layer undirected graphical model consisting of layers of visible and hidden units. Compared to the previous models, RBMMMDA could obtain not only new miRNA-disease associations but also corresponding association types. However, it is still too difficult to learn the complex parameters.

In this study, we developed an effective computational model of Matrix Completion for MiRNA-Disease Association prediction model (MCMDA) using matrix completion algorithm based on the known miRNA-disease associations to predict the potential miRNA-disease associations. Compared to the previous computational models, MCMDA predicts the miRNA-disease associations by using the matrix completion algorithm, which is of high efficiency to update the low-rank miRNA-disease matrix. Besides, negative associations which are required in some previous computational models are not needed in MCMDA. To evaluate the effectiveness of MCMDA, global and local LOOCV as well as 5-fold cross validation were introduced. The AUCs of global and local LOOCV were respectively 0.8749 and 0.7718, and the model obtained the average AUC of 0.8767+/-0.0011 on 5-fold cross validation. Besides, the top 10 and top 50 miRNAs related to colon neoplasms, kidney neoplasms, lymphoma and prostate neoplasms obtained by MCMDA were examined in dbDEMC [44] and miR2Disease [45] database. As a result, 84%, 86%, 78% and 90% of the top 50 potential miRNAs for these four complex diseases were respectively confirmed by recent experimental discoveries. Thus, it proves that MCMDA is effective in predicting potential miRNA-disease associations and it has significant advantages over the previous methods although MCMDA only depends on known miRNA-disease associations.

RESULTS

Performance evaluation

We used global and local LOOCV as well as 5-fold cross validation based on the known miRNA-disease associations in HMDD database to evaluate the performance of MCMDA. Meanwhile, MCMDA were compared with three previous classical computational methods: WBSMDA [16], RLSMDA [15] and HDMP [38]. In LOOCV evaluation, each known association in the database was regarded as the test sample in turn while the other known associations were regarded as training samples. The miRNA-diseases without known association evidences were considered as candidate samples. The scores of all miRNA-disease pairs could be obtained after MCMDA was implemented. In global LOOCV, the score of the test sample was compared with the scores of all the candidate samples while in local LOOCV, the test sample was merely compared with the scores of the candidate samples which included the particular disease in the test sample. In 5-fold cross validation, the known miRNA-disease associations were randomly divided into five disjoint parts. Each time, one part was picked out as test samples and the other four parts were treated as training samples. Still, the miRNA-disease pairs without known association evidences were regarded as candidate samples. Then, the score of each test sample were compared with the scores of all the candidate samples, respectively. This procedure was repeated five times until each known association was used as test sample and its score was compared with the scores of the candidate samples. Those test samples whose ranks exceeded the given threshold were considered to predict the miRNA-disease associations correctly.

Finally, we drew a receiver operating characteristics curve (ROC) to compare MCMDA with all the previous methods. In this curve, the true positive rate (TPR, sensitivity) and false positive rate (FPR, 1-specificity) were plotted [46]. Sensitivity represents the percentage

of miRNA-disease test samples whose ranks exceeded the given threshold while specificity represents the percentage of negative miRNA-disease associations whose ranks were lower than the threshold [47]. The area under the ROC curve (AUC) was calculated to evaluate the accuracy of MCMDA. If AUC=1, MCMDA proves to be a prefect performance. AUC of 0.5 means that the method merely has a random prediction performance. As a result, the AUCs of MCMDA, WBSMDA, RLSMDA and HDMP were 0.8749, 0.8030, 0.8426, and 0.8366, respectively in global LOOCV. For local LOOCV, MCMDA, WBSMDA, RLSMDA and HDMP acquired AUCs of 0.7718, 0.8030, 0.8031 and 0.6953, respectively. The average AUCs of MCMDA, WBSMDA, RLSMDA, HDMP were 0.8767/-0.0011, 0.8185/-0.0009, 0.8569/-0.0020 and 0.8342+/-0.0010, respectively in 5-fold cross validation (See Figure 1). All in all, MCMDA turns out to be more effective in predicting potential miRNA-disease associations compared with the previous methods, especially considering that MCMDA merely depends on the known miRNA-disease associations in the database.

Case studies

Furthermore, case studies of four significant diseases related to human health were implemented to practically evaluate the prediction accuracy of MCMDA. The top 10 and top 50 predicted miRNAs related with these diseases were examined by another two miRNA-disease databases, dbDEMC [44] and miR2Disease [45].

Colon Neoplasms is a malignant cancer which is commonly found in the boundary of rectum and sigmoid colon [48]. It is the third most common cancer and the third leading cause of cancer death for both men and women in the United States [49]. However, early patients of colon neoplasms only suffer from subtle symptoms [50], making the disease difficult to be detected. To make things worse, it is reported that its occurrence rate has an increasing trend these years [51]. Thus, it is urgent to predict the potential miRNAs related to colon neoplasms. With the help of the modern iatrology, many miRNAs have been confirmed to be correlated with colon neoplasms. For instance, miRNA-145 targets the insulin receptor substrate-1 and thus inhibits the growth of colon cancer cells [52]. Besides, miRNA-126, which is frequently lost in colon neoplasms cells, has the function of suppressing the growth of neoplastic cells by targeting phosphatidylinositol 3-kinase signaling [53]. MCMDA was implemented to predict the top 50 miRNAs associated with colon neoplasms. Therefore, 9 of the top 10 and 42 of the top 50 predicted miRNAs associated with colon neoplasms were verified by dbDEMC and miR2Disease database (See Table 1).

Kidney neoplasms, also known as renal cancer, is a cancer starting in the cells of kidney that includes many different types [54]. The two most common types of kidney
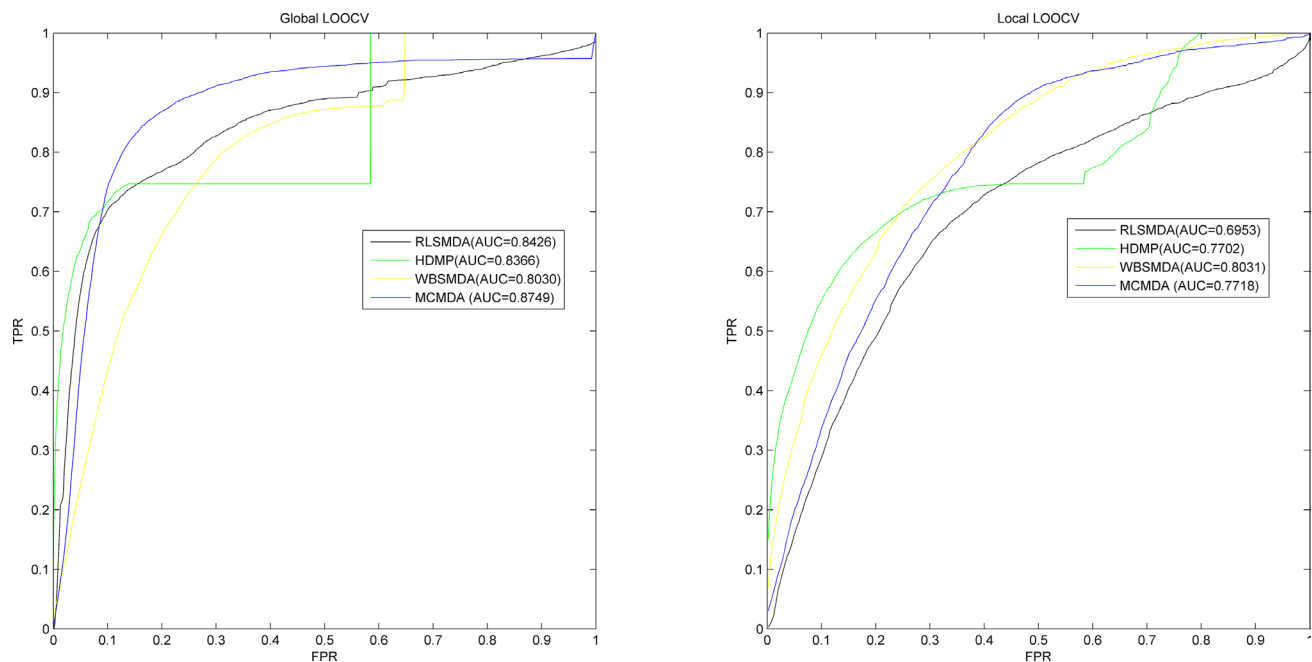

Figure 1: Performance evaluation comparison between MCMDA and three previous prediction models (RLSMDA, HDMP, WBSMDA) in terms of ROC curve and AUC based on global LOOCV and local LOOCV tested by known miRNA-disease associations in the HMDD database. MCMDA achieved AUC of 0.8749 in global LOOCV and 0.7718 in local LOOCV. Thus, the performance of MCMDA is almost better than all the previous models in some degree and it proves to be effective in predicting the potential miRNA-disease associations.

Table 1: Prediction of the top 50 predicted miRNAs associated with colon neoplasms based on known associations in HMDD database

miRNA	Evidence	miRNA	Evidence
hsa-mir-146a	dbdemc	hsa-mir-196a	dbdemc;miR2Disease
hsa-mir-155	dbdemc;miR2Disease	hsa-mir-29c	dbdemc
hsa-mir-122	unconfirmed	hsa-mir-223	dbdemc;miR2Disease
hsa-mir-21	dbdemc;miR2Disease	hsa-mir-143	dbdemc;miR2Disease
hsa-mir-34a	dbdemc;miR2Disease	hsa-let-7a	unconfirmed
hsa-mir-221	dbdemc;miR2Disease	hsa-mir-195	dbdemc;miR2Disease
hsa-mir-16	dbdemc	hsa-mir-200b	dbdemc
hsa-mir-125b	dbdemc	hsa-mir-214	dbdemc
hsa-mir-29a	dbdemc;miR2Disease	hsa-mir-106b	dbdemc;miR2Disease
hsa-mir-29b	dbdemc;miR2Disease	hsa-mir-23a	miR2Disease
hsa-mir-15a	dbdemc	hsa-mir-142	unconfirmed
hsa-mir-133a	dbdemc;miR2Disease	hsa-mir-31	dbdemc;miR2Disease
hsa-mir-222	dbdemc	hsa-mir-34c	miR2Disease
hsa-mir-20a	dbdemc;miR2Disease	hsa-mir-141	dbdemc;miR2Disease
hsa-mir-199a	unconfirmed	hsa-mir-148a	dbdemc
hsa-mir-26a	dbdemc;miR2Disease	hsa-mir-182	dbdemc;miR2Disease
hsa-mir-1	dbdemc;miR2Disease	hsa-mir-200a	unconfirmed
hsa-mir-19b	dbdemc;miR2Disease	hsa-let-7c	dbdemc
hsa-mir-19a	dbdemc;miR2Disease	hsa-mir-101	unconfirmed
hsa-mir-15b	miR2Disease	hsa-mir-192	dbdemc;miR2Disease
hsa-mir-18a	miR2Disease	hsa-mir-181a	dbdemc;miR2Disease
hsa-mir-92a	dbdemc	hsa-mir-9	dbdemc;miR2Disease
hsa-mir-206	unconfirmed	hsa-mir-133b	dbdemc;miR2Disease
hsa-mir-30b	dbdemc;miR2Disease	hsa-mir-34b	dbdemc;miR2Disease
hsa-mir-150	dbdemc;miR2Disease	hsa-mir-183	dbdemc;miR2Disease

The first column records top 1-25 related miRNAs. The second column records the top 26-50 related miRNAs.

cancer are renal cell carcinoma (RCC) and transitional cell carcinoma (TCC, also known as urothelial cell carcinoma) of the renal pelvis [55]. The most common symptoms of kidney neoplasms patients are pains in the lumbar and hematuria [56]. Many existing kidney neoplasm-related miRNAs have been reported based on recent biological experiments. For example, the common target ACVR2B of five miRNAs (miRNA-192, miRNA-194, miRNA-215, miRNA-200c and miRNA-141) is strongly expressed in renal childhood neoplasms [57]. In addition, miRNA-23b, by targeting proline oxidase, a novel tumor suppressor protein, could function as an oncogene in renal cancer [58]. Thus, the decreasing miRNA-23b expression may prove to be an effective way of inhibiting kidney tumor growth [58]. Based on MCMDA, 7 of the top 10 potential miRNAs associated with kidney neoplasms were confirmed by deDEMC and miR2Disease database while 43 were verified of the top 50 (See Table 2).

Lymphoma is a malignant tumor originating in the lymphatic hematopoietic system [59] which consists of two categories: non-Hodgkinlymphoma (NHL) and Hodgkin'slymphoma (HL) [60]. Lymphoma is thought

Table 2: Prediction of the top 50 predicted miRNAs associated with kidney neoplasms based on known associations in HMDD database

miRNA	Evidence	miRNA	Evidence
hsa-mir-155	dbdemc	hsa-mir-92a	unconfirmed
hsa-mir-146a	dbdemc	hsa-mir-195	dbdemc
hsa-mir-122	dbdemc;miR2Disease	hsa-mir-126	dbdemc;miR2Disease
hsa-mir-34a	dbdemc	hsa-mir-29c	dbdemc;miR2Disease
hsa-mir-221	unconfirmed	hsa-mir-23a	dbdemc
hsa-mir-16	dbdemc	hsa-mir-143	dbdemc
hsa-mir-125b	unconfirmed	hsa-mir-223	dbdemc
hsa-mir-29a	dbdemc;miR2Disease	hsa-mir-214	dbdemc;miR2Disease
hsa-mir-133a	unconfirmed	hsa-let-7a	dbdemc
hsa-mir-29b	dbdemc;miR2Disease	hsa-mir-148a	dbdemc
hsa-mir-145	dbdemc	hsa-mir-200b	dbdemc;miR2Disease
hsa-mir-26a	dbdemc;miR2Disease	hsa-mir-31	dbdemc
hsa-mir-199a	dbdemc;miR2Disease	hsa-mir-142	unconfirmed
hsa-mir-222	dbdemc	hsa-mir-106b	dbdemc;miR2Disease
hsa-mir-1	dbdemc	hsa-mir-34c	dbdemc
hsa-mir-15b	dbdemc	hsa-mir-182	dbdemc;miR2Disease
hsa-mir-20a	dbdemc;miR2Disease	hsa-mir-200a	dbdemc
hsa-mir-17	dbdemc;miR2Disease	hsa-mir-101	dbdemc;miR2Disease
hsa-mir-30b	dbdemc	hsa-let-7c	dbdemc
hsa-mir-206	dbdemc	hsa-mir-181a	dbdemc
hsa-mir-19a	dbdemc	hsa-mir-9	dbdemc
hsa-mir-196a	dbdemc	hsa-mir-34b	dbdemc
hsa-mir-19b	dbdemc;miR2Disease	hsa-mir-183	dbdemc
hsa-mir-18a	dbdemc	hsa-mir-133b	unconfirmed
hsa-mir-150	dbdemc;miR2Disease	hsa-let-7b	unconfirmed

The first column records top 1-25 related miRNAs. The second column records the top 26-50 related miRNAs.

to be associated with gene mutations, as well as viruses, pathogens, radiation, chemical drugs, autoimmune diseases, etc. [61]. For example, re-expression of miRNA-150 induces EBV-positive Burkitt lymphoma differentiation by modulating c-Myb *in vitro* [62]. Besides, the expressions of miRNA-21 and miRNA-210 in plasma of previously untreated lymphoma patient group were higher than those of the patients treated for 6 or more courses [63]. MCMDA model predicts the top 10 and top 50 miRNAs related to lymphoma. As a result, 9 of the top 10 and 39 of the top 50 potential miRNAs were confirmed in the deDEMC and miR2Disease database (See Table 3).

Prostate neoplasms is a malignant tumor which originates in the epithelial cells of prostate [64]. Factors that increase the risk of prostate neoplasms include older age, a family history of the disease, race and a diet high in processed meat, red meat or milk products or low in certain vegetables [65]. Up to now, lots of miRNAs have been discovered to be associated with prostate neoplasms. For instance, the proto-oncogene ERG is a target of miRNA-145 in prostate cancer [66]. MCMDA predicts the top 10 and top 50 potential miRNAs which are associated with prostate neoplasms. As a consequence, 9 of the top 10 and 45 of the top 50

Table 3: Prediction of the top 50 predicted miRNAs associated with lymphoma based on known associations in HMDD database

miRNA	Evidence	miRNA	Evidence
hsa-mir-30b	dbdemc	hsa-mir-208a	unconfirmed
hsa-mir-148a	dbdemc	hsa-mir-26b	dbdemc
hsa-mir-373	dbdemc	hsa-mir-143	unconfirmed
hsa-mir-196a	dbdemc	hsa-mir-9	dbdemc
hsa-mir-23a	dbdemc	hsa-let-7b	dbdemc
hsa-mir-206	dbdemc	hsa-mir-96	dbdemc
hsa-mir-195	dbdemc	hsa-let-7d	dbdemc
hsa-mir-372	unconfirmed	hsa-mir-93	dbdemc
hsa-mir-199a	dbdemc	hsa-mir-483	unconfirmed
hsa-mir-15b	dbdemc	hsa-mir-371a	unconfirmed
hsa-mir-34c	unconfirmed	hsa-let-7e	dbdemc;miR2Disease
hsa-mir-34b	dbdemc	hsa-mir-7	dbdemc
hsa-mir-183	dbdemc	hsa-mir-223	dbdemc
hsa-mir-132	dbdemc	hsa-mir-106a	dbdemc;miR2Disease
hsa-mir-214	dbdemc	hsa-mir-205	dbdemc
hsa-mir-182	dbdemc	hsa-mir-222	dbdemc
hsa-mir-31	unconfirmed	hsa-mir-335	dbdemc
hsa-mir-133a	dbdemc	hsa-mir-27a	dbdemc
hsa-mir-212	dbdemc	hsa-mir-181c	dbdemc
hsa-mir-141	dbdemc	hsa-mir-224	dbdemc
hsa-mir-142	unconfirmed	hsa-mir-27b	dbdemc
hsa-mir-192	dbdemc	hsa-mir-30a	dbdemc
hsa-mir-429	unconfirmed	hsa-mir-370	unconfirmed
hsa-mir-451a	unconfirmed	hsa-mir-1	dbdemc
hsa-mir-106b	dbdemc	hsa-let-7g	dbdemc

The first column records top 1-25 related miRNAs. The second column records the top 26-50 related miRNAs.

predicted miRNAs were confirmed in the dbDEMC and miR2Disease database (See Table 4).

The result of case studies on the four aforementioned human diseases illustrates that MCMDA achieves excellent prediction performance. Moreover, we prioritized the potential miRNAs associated with all the human diseases in HMDD database (See Supplementary Table 1). We hope that the predictions of MCMDA can be verified in future scientific researches.

DISCUSSION

Nowadays, researchers propose several computational methods to predict the potential associations between miRNAs and diseases because computational models could select the most promising miRNAs related to human diseases and are less expensive than the traditional experimental methods. In order to predict potential miRNA-disease associations, we developed a computational model of MCMDA by analyzing the known miRNA-disease associations and implementing the matrix completion algorithm to get the association score of each miRNA-disease pair. MCMDA obtained excellent prediction performances based on LOOCV and 5-fold cross validation. In addition, the predicted miRNAs associated with four important human diseases: colon neoplasms, kidney neoplasms, lymphoma and prostate neoplasms, were verified by the experimental literatures

Table 4: Prediction of the top 50 predicted miRNAs associated with prostate neoplasms based on known associations in HMDD database

miRNA	Evidence	miRNA	Evidence
hsa-mir-146a	miR2Disease	hsa-mir-150	dbdemc
hsa-mir-122	unconfirmed	hsa-mir-126	dbdemc;miR2Disease
hsa-mir-155	dbdemc	hsa-mir-195	dbdemc;miR2Disease
hsa-mir-21	dbdemc;miR2Disease	hsa-mir-29c	dbdemc
hsa-mir-34a	dbdemc;miR2Disease	hsa-mir-223	dbdemc;miR2Disease
hsa-mir-16	dbdemc;miR2Disease	hsa-mir-143	dbdemc;miR2Disease
hsa-mir-221	dbdemc;miR2Disease	hsa-mir-23a	dbdemc;miR2Disease
hsa-mir-29a	dbdemc	hsa-let-7a	dbdemc;miR2Disease
hsa-mir-133a	dbdemc	hsa-mir-200b	unconfirmed
hsa-mir-29b	dbdemc;miR2Disease	hsa-mir-214	dbdemc;miR2Disease
hsa-mir-15a	dbdemc;miR2Disease	hsa-mir-148a	miR2Disease
hsa-mir-26a	dbdemc;miR2Disease	hsa-mir-106b	dbdemc
hsa-mir-222	dbdemc;miR2Disease	hsa-mir-34c	dbdemc
hsa-mir-199a	dbdemc;miR2Disease	hsa-mir-142	unconfirmed
hsa-mir-1	dbdemc	hsa-mir-31	dbdemc;miR2Disease
hsa-mir-20a	miR2Disease	hsa-mir-141	miR2Disease
hsa-mir-17	miR2Disease	hsa-mir-182	dbdemc;miR2Disease
hsa-mir-15b	dbdemc	hsa-mir-200a	dbdemc
hsa-mir-19a	dbdemc	hsa-mir-101	dbdemc;miR2Disease
hsa-mir-19b	dbdemc;miR2Disease	hsa-let-7c	dbdemc;miR2Disease
hsa-mir-206	dbdemc	hsa-mir-192	dbdemc
hsa-mir-30b	dbdemc;miR2Disease	hsa-mir-181a	dbdemc;miR2Disease
hsa-mir-18a	unconfirmed	hsa-mir-9	dbdemc
hsa-mir-196a	dbdemc	hsa-mir-34b	dbdemc
hsa-mir-92a	unconfirmed	hsa-mir-133b	dbdemc

The first column records top 1-25 related miRNAs. The second column records the top 26-50 related miRNAs.

in dbDEMC and miR2Disease database. The results from cross validation and case studies indicated that MCMDA was effective in predicting potential miRNA-disease associations although it only depends on known miRNA-disease associations.

The reasons why MCMDA achieved excellent performances are as follows. Firstly, MCMDA predicts the miRNA-disease associations by using the matrix completion algorithm based on the observation that the miRNA-disease matrix is low-rank. MCMDA fills the candidate samples without known associations with 0 and then iteratively updates them with the predictive scores. Besides, MCMDA is based on the known miRNA-disease associations in HMDD database. Plenty of known

associations guarantee the efficiency of the predictions in MCMDA. Finally, negative associations which are required in some previous models are not needed in MCMDA.

Yet, there still exist several limitations in MCMDA. Firstly, MCMDA method is based on the known miRNA-disease associations, which means it cannot predict the potential miRNAs associated with the new diseases without any known related miRNAs and potential diseases associated with new miRNAs. Besides, there is no powerful method to find the optimal parameters for MCMDA. Finally, the current miRNA-disease associations are insufficient. To be specific, there are merely 5430 known miRNA-disease associations within
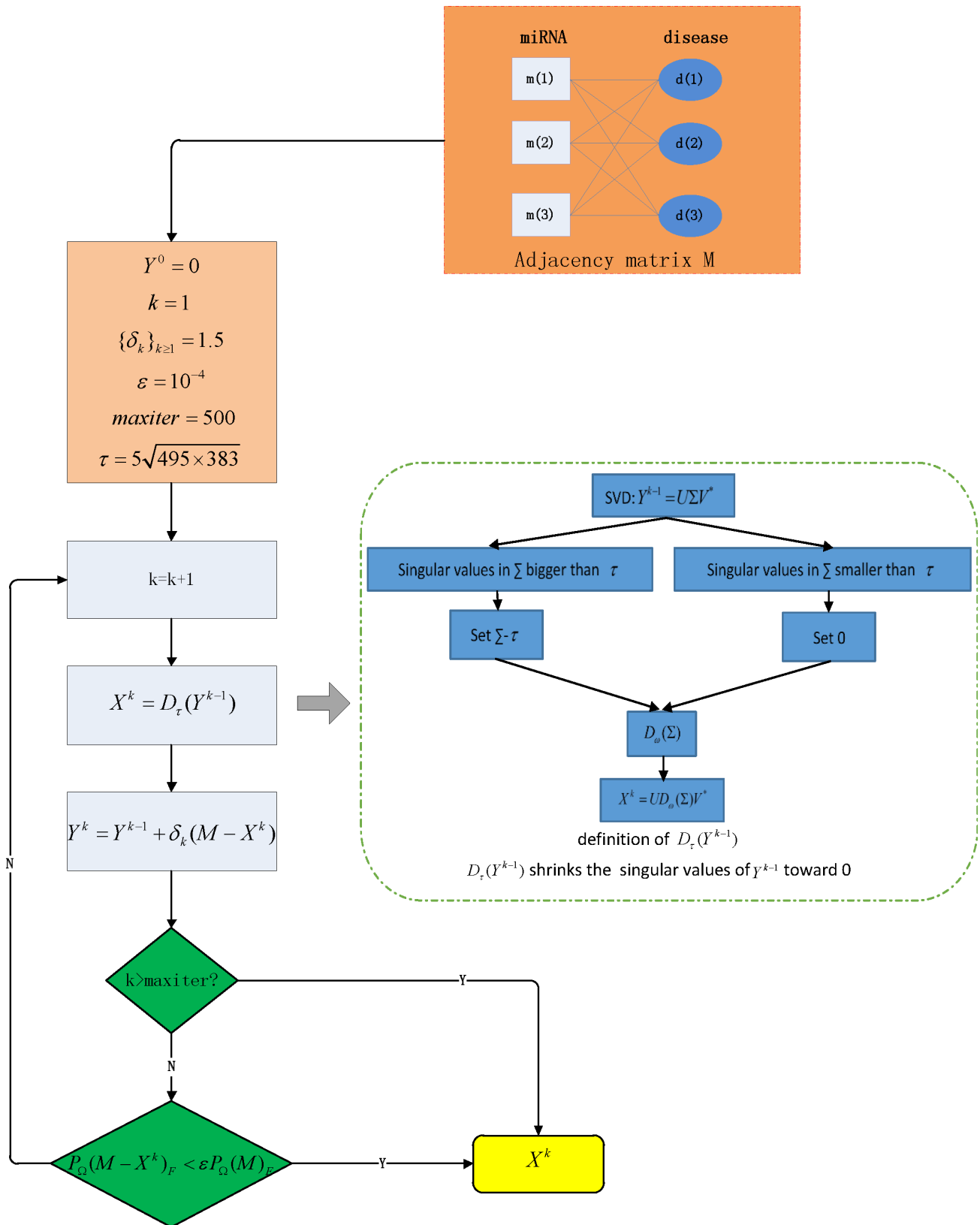
Figure 2: Flowchart of MCMDA model to predict the potential miRNA-disease associations based on the known associations in HMDD database.

the possible exploration spaces of 495 miRNAs and 383 diseases. The more known associations are confirmed in the future, the more accurate MCMDA model can become.

MATERIALS AND METHODS

Human miRNA-disease associations

The known miRNA-disease associations were downloaded from HMDD v2.0 database [67] which consisted of 5430 known miRNA-disease associations, 495 miRNAs, and 383 diseases. We furthermore constructed an adjacency matrix M to represent known miRNA-disease associations. For instance, if miRNA $m(i)$ is reported to be associated with disease $d(j)$ in the database, the value of $M(i, j)$ is 1 and otherwise 0. Ω denotes the set of all the known associations in matrix M which means (i, j) Ω if $m(i)$ is associated with $d(j)$. n_m represents the number of miRNAs in HMDD database and n_d represents the number of diseases.

MCMDA

We developed MCMDA based on the known miRNA-disease associations in HMDD database to predict the potential associations (See Figure 2). MCMDA uses the singular value thresholding (SVT) algorithm to accomplish the matrix completion procedure. First, the miRNA-disease association matrix M was obtained according to known miRNA-disease associations. Here, all the known associations between miRNAs and diseases in HMDD database are used as training samples.

The matrix completion algorithm is iterative and a $n_m \times n_d$ prediction matrix X^k (k denotes the iteration times) can be obtained in each iteration. When MCMDA ends, the matrix X^n (n denotes the ultimate iteration times) is obtained which records the scores of all the possible miRNA-disease pairs. To ensure that the scores of known associations in X^n are close to those in M, the following optimization problem needs to be solved.

$$\min_X f_t(X)$$
$$s.t. P_\Omega(X) = P_\Omega(M) \tag{1}$$

where X is a $n_m \times n_d$ candidate solution matrix with scores of all the unknown miRNA-disease samples, P_Ω is the orthogonal projector onto the span of matrices vanishing outside of Ω so that the (i,j) th component of $P_\Omega(X)$ is equal to $X(i, j)$ if (i, j) Ω or zero otherwise. $f_\tau(X)$ is a nonlinear function of X which can be written as the following form.

$$\min_X \tau \|X\|_* + \frac{1}{2} \|X\|_F^2$$
$$s.t. P_\Omega(X) = P_\Omega(M) \tag{2}$$

where $\|X\|_*$ is the nuclear form of the matrix X which is the sum of the singular values of X, $\|X\|_F$ denotes the Frobernius form of X which is $\sqrt{\sum_{i=1}^{n_m} \sum_{j=1}^{n_d} X(i, j)^2}$, t is a thresholding which will be introduced later.

According to [68], problem (2) can be optimized using the Lagrangian multiplier method. Specifically, we introduce a Lagrangian multiplier Y and get the Lagrangian function as below:

$$L(X, Y) = f_\tau(X) + < Y, P_\Omega(M) - P_\Omega(X) > \tag{3}$$

The singular value decomposition (SVD) of matrix X with rank r, which represents the number of singular values of matrix X, is needed in matrix completion algorithm.

$$X = U \Sigma V^*, \Sigma = diag(\{\sigma_i\}_{1 \le i \le r}) \tag{4}$$

where U and V are $n_m \times r$ and $r \times n_d$ matrices. $S = diag(\{s_i\}_{1 \le i \le r})$ means that Σ is a $r \times r$ diagonal matrix with positive singular values $\{s_i\}_{1 \le i \le r}$ on its main diagonal. For $t \ge 0$, we introduce an operator D_τ defined as follows:

$$D_t(X) = U D_w(S) V^*, D_w(S) = diag(\{s_i - t\}_+) \tag{5}$$

where $\{\sigma_i - \tau\}_+$ is the positive part of $\{s_i - t\}$. In other words, $\{s_i - t\}_+$ is equal to $s_i - t$ if $s_i - t \ge 0$ or 0 otherwise and it effectively shrinks the singular values of X toward 0. The value of t is $5\sqrt{n_m \times n_d}$ according to the previous research of matrix completion algorithm [69].

There are two key steps which are special instances of Uzawa's algorithm [70] to find a saddle point of (3) in each iteration. We introduce $\{Y^0, Y^1, ..., Y^{n-1}, Y^n\}$ which are a series of $n_m \times n_d$ matrices to record the intermediate scores of matrices $\{X^1..X^n\}$. First, update X with Y:

$$X^k = D_\tau(Y^{k-1}) \tag{6}$$

Then, update Y with X:

$$Y^k = Y^{k-1} + \delta_k(M - X^k) \tag{7}$$

where Y^0 is a zero matrix [71] and $\{d_k\}_{k \ge 1}$ is the step size. It is usually thought that the iteration can converge to an unique solution when $0 < d < 2$ [72], specifically, we empirically set the value of $\{d_k\}_{k \ge 1} = 1.5$ according to the excellent performance in previous model [73]. MCMDA applies K.K.T conditions as the stopping criteria which are checked in each iteration to makes sure the scores of the known associations in the prediction matrix are close enough to the original matrix M:

$$P_\Omega(M - X^k)_F < \varepsilon P_\Omega(M)_F \tag{8}$$

where $\in$ is a stopping tolerance, the value is 10^{-4} since it proved to be appropriate in restricting the iteration times

in previous algorithm [71]. If the stopping criteria is met, MCMDA stops iteration immediately and the ultimate matrix X^n is obtained. Finally, a parameter *maxiter* is set which restricts the max iteration times and avoids the infinite loop. Specifically, *maxiter* is set 500 to ensure that the ultimate matrix has reliable predicted scores. Based on the method mentioned above, the ultimate matrix X^n is obtained by above calculation process which can be utilized to predict the potential miRNA-disease associations.

ACKNOWLEDGMENTS

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

REFERENCES

1. Llave C, Xie Z, Kasschau KD, Carrington JC. Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. Science. 2002; 297:2053-2056.

2. Eulalio A, Huntzinger E, Izaurralde E. Getting to the Root of miRNA-Mediated Gene Silencing. Cell. 2008; 132:9-14.

3. Telese F, Gamliel A, Skowronska-Krawczyk D, Garcia-Bassets I, Rosenfeld M. "Seq-ing" Insights into the Epigenetics of Neuronal Gene Regulation. Neuron. 2013; 77:606-623.

4. Lu J, Clark AG. Impact of microRNA regulation on variation in human gene expression. Genome Res. 2012; 22:1243-1254.

5. Zhu L, Zhao J, Wang J, Hu C, Peng J, Luo R, Zhou C, Liu J, Lin J, Jin Y. MicroRNAs Are Involved in the Regulation of Ovary Development in the Pathogenic Blood Fluke Schistosoma japonicum. Plos Pathogens. 2016; 12:e1005423.

6. Fernando TR, Rodriguez-Malave NI, Rao DS. MicroRNAs in B cell development and malignancy. J Hematol Oncol. 2012; 5:7.

7. Lizé M, Pilarski S, Dobbelstein M. E2F1-inducible microRNA 449a/b suppresses cell proliferation and promotes apoptosis. Cell Death Differ. 2009; 17:452-458.

8. Kim YK. Extracellular microRNAs as Biomarkers in Human Disease. Chonnam Med J. 2015; 51:51-57.

9. Alaimo S, Giugno R, Pulvirenti A. ncPred: ncRNA-Disease Association Prediction through Tripartite Network-Based Inference. Front Bioeng Biotechnol. 2014; 2:71.

10. Li JM, Zhao RH, Li ST, Xie CX, Jiang HH, Ding WJ, Du P, Chen W, Yang M, Cui L. Down-regulation of fecal miR-143 and miR-145 as potential markers for colorectal cancer. Saudi Med J. 2012; 33:24-29.

11. Stahlhut Espinosa CE, Slack FJ. The role of microRNAs in cancer. Yale J Biol Med. 2006; 79: 131–140.

12. Takamizawa J, Konishi H, Yanagisawa K, Tomida S, Osada H, Endoh H, Harano T, Yatabe Y, Nagino M, Nimura Y. Reduced expression of the let-7 microRNAs in human lung cancers in association with shortened postoperative survival. Cancer Res. 2004; 64:3753-3756.

13. Yanaihara N, Caplen N, Bowman E, Seike M, Kumamoto K, Yi M, Stephens RM, Okamoto A, Yokota J, Tanaka T. Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. Cancer Cell. 2006; 9:189-198.

14. Inoue T, Iinuma H, Ogawa E, Inaba T, Fukushima R. Clinicopathological and prognostic significance of microRNA-107 and its relationship to DICER1 mRNA expression in gastric cancer. Oncol Rep. 2012; 27:1759-1764.

15. Chen X, Yan GY. Semi-supervised learning for potential human microRNA-disease associations inference. Sci Rep. 2014; 4:5501.

16. Chen X, Yan CC, Zhang X, You ZH, Deng L, Liu Y, Zhang Y, Dai Q. WBSMDA: Within and Between Score for MiRNA-Disease Association prediction. Sci Rep. 2016; 6:21106.

17. 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.

18. Chen X. Predicting lncRNA-disease associations and constructing lncRNA functional similarity network based on the information of miRNA. Sci Rep. 2015; 5:13186.

19. Chen X, Yan CC, Zhang X, Li Z, Deng L, Zhang Y, Dai Q. RBMMMDA: predicting multiple types of disease-microRNA associations. Sci Rep. 2015; 5:13877.

20. Chen X, Yan CC, Luo C, Ji W, Zhang Y, Dai Q. Constructing lncRNA functional similarity network based on lncRNA-disease associations and disease semantic similarity. Sci Rep. 2015; 5:11338.

21. Chen X, Yan GY. Novel human lncRNA-disease association inference based on lncRNA expression profiles. Bioinformatics. 2013; 29:2617-2624.

22. Chen X, You Z, Yan G, Gong D. IRWRLDA: Improved Random Walk with Restart for LncRNA-Disease Association prediction. Oncotarget. 2016; 7:57919-57931. doi: 10.18632/oncotarget.11141.

23. Chen X, Yan CC, Zhang X, You Z-H. Long non-coding RNAs and complex diseases: from experimental results to computational models. Briefings in Bioinformatics. 2016:bbw060.

24. Chen X, Liu MX, Cui QH, Yan GY. Prediction of Disease-Related Interactions between MicroRNAs and Environmental Factors Based on a Semi-Supervised Classifier. PLoS One. 2012; 7:e43425.

25. Chen X. KATZLDA: KATZ measure for the lncRNA-disease association prediction. Sci Rep. 2015; 5:16840.

26. Chen X. miREFRWR: a novel disease-related microRNA-environmental factor interactions prediction method. Mol Biosyst. 2016; 12:624-633.

27. Huang YA, You ZH, Xing C, Chan K, Xin L. Sequence-based prediction of protein-protein interactions using weighted sparse representation model combined with global encoding. Bmc Bioinformatics. 2016; 17:184.

28. Huang YA, Chen X, You ZH, Huang DS, Chan KC. ILNCSIM: improved lncRNA functional similarity calculation model. Oncotarget. 2016; 7:25902-25914. doi: 10.18632/oncotarget.8296.

29. Chen X, Huang Y, Wang X, You Z, Chan K. FMLNCSIM: fuzzy measure-based lncRNA functional similarity calculation model. Oncotarget. 2016; 7:45948-45958. doi: 10.18632/oncotarget.10008.

30. Wong L, You Z-H, Ming Z, Li J, Chen X, Huang Y-A. Detection of interactions between proteins through rotation forest and local phase quantization descriptors. International journal of molecular sciences. 2016; 17:21.

31. Lan W, Li M, Zhao K, Jin L, Wu FX, Yi P, Wang J. LDAP: a web server for lncRNA-disease association prediction. Bioinformatics. 2016.

32. Pasquier C, Gardès J. Prediction of miRNA-disease associations with a vector space model. Sci Rep. 2016; 6:27036.

33. Bandyopadhyay S, Mitra R, Maulik U, Zhang MQ. Development of the human cancer microRNA network. Silence. 2010; 1:6.

34. 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.

35. Shi H, Xu J, Zhang G, Xu L, Li C, Li W, Zheng Z, Wei J, Zheng G, Xia L. 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.

36. Mørk S, Pletscherfrankild S, Caro AP, Gorodkin J, Jensen LJ. Protein-driven inference of miRNA–disease associations. Bioinformatics. 2014; 30:392-397.

37. Xu C, Ping Y, Li X, Zhao H, Wang L, Fan H, Xiao Y, Li X. Prioritizing candidate disease miRNAs by integrating phenotype associations of multiple diseases with matched miRNA and mRNA expression profiles. Mol Biosyst. 2014; 10:2800-2809.

38. Xuan P, Han K, Guo M, Guo Y, Li J, Ding J, Liu Y, Dai Q, Li J, Teng Z. Prediction of microRNAs Associated with Human Diseases Based on Weighted k Most Similar Neighbors. PLoS One. 2013; 8:e70204.

39. Xuan P, Han K, Guo Y, Li J, Li X, Zhong Y, Zhang Z, Ding J. Prediction of potential disease-associated microRNAs based on random walk. Bioinformatics. 2015; 31:1805-1815.

40. Zeng X, Ding N, Rodríguezpatón A, Lin Z, Ju Y. Prediction of MicroRNA–disease Associations by Matrix Completion. Current Proteomics. 2016; 13:151-157.

41. Chen X, Liu MX, Yan GY. RWRMDA: predicting novel human microRNA-disease associations. Mol Biosyst. 2012; 8:2792-2798.

42. Chen X, Yan C, Zhang X, You Z, Huang Y, Yan G. HGIMDA: Heterogeneous Graph Inference for MiRNA-Disease Association prediction. Oncotarget. 2016; 7:65257-65269. doi: 10.18632/oncotarget.11251.

43. Xu J, Li CX, Lv JY, Li YS, Xiao Y, Shao TT, Huo X, Li X, Zou Y, Han QL. 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.

44. 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 genomics. 2010; 11:S5.

45. 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.

46. 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.

47. Oved K, Morag A, Pasmanikchor M, Oronkarni V, Shomron N, Rehavi M, Stingl JC, Gurwitz D. Genome-wide miRNA expression profiling of human lymphoblastoid cell lines identifies tentative SSRI antidepressant response biomarkers. Pharmacogenomics. 2012; 13:1129-1139.

48. Phipps AI, Lindor NM, Jenkins MA, Baron JA, Win AK, Gallinger S, Gryfe R, Newcomb PA. Colon and rectal cancer survival by tumor location and microsatellite instability: the Colon Cancer Family Registry. Dis Colon Rectum. 2013; 56:937-944.

49. Liu F, Yuan D, Wei Y, Wang W, Yan L, Wen T, Xu M, Yang J, Li B. Systematic review and meta-analysis of the relationship between EPHX1 polymorphisms and colorectal cancer risk. PLoS One. 2012; 7:e43821.

50. Pita-Fernández S, Pértega-Díaz S, López-Calviño B, Seoane-Pillado T, Gago-García E, Seijo-Bestilleiro R, González-Santamaría P, Pazos-Sierra A. Diagnostic and treatment delay, quality of life and satisfaction with care in colorectal cancer patients: a study protocol. Health Qual Life Outcomes. 2013; 11:117.

51. Chong VH, Abdullah MS, Telisinghe PU, Jalihal A. Colorectal cancer: incidence and trend in Brunei Darussalam. Singapore Med J. 2009; 50:1085-1089.

52. Shi B, Sepplorenzino L, Prisco M, Linsley P, Deangelis T, Baserga R. Micro RNA 145 targets the insulin receptor substrate-1 and inhibits the growth of colon cancer cells. J Biol Chem. 2007; 282:32582-32590.

53. Guo C, Sah JF, Beard L, Willson JKV, Markowitz SD, Guda K. The noncoding RNA, miR-126, suppresses the growth of neoplastic cells by targeting phosphatidylinositol 3-kinase signaling and is frequently lost in colon cancers. Genes Chromosomes Cancer. 2008; 47:939-946.

54. Manojlović D, Elaković D, Lj M. [Therapeutic value of transcatheter embolization in malignant tumors of the renal parenchyma]. Srp Arh Celok Lek. 1986; 114:631-637.

55. Takehara K, Nishikido M, Koga S, Miyata Y, Harada T, Tamaru N, Kanetake H. Multifocal transitional cell carcinoma associated with renal cell carcinoma in a patient on long-term haemodialysis. Nephrol Dial Transplant. 2002; 17:1692-1694.

56. Duque JLF, Loughlin KR, O'Leary MP, Kumar S, Richie JP. Partial nephrectomy: alternative treatment for selected patients with renal cell carcinoma. Urology. 1998; 52:584-590.

57. Senanayake U, Das S, Vesely P, Alzoughbi W, Fröhlich LF, Chowdhury P, Leuschner I, Hoefler G, Guertl B. miR-192, miR-194, miR-215, miR-200c and miR-141 are downregulated and their common target ACVR2B is strongly expressed in renal childhood neoplasms. Carcinogenesis. 2012; 33:1014-1021.

58. Liu W, Zabirnyk OH, Shiao YH, Nickerson ML, Khalil S, Anderson LM, Perantoni AO, Phang JM. miR-23b targets proline oxidase, a novel tumor suppressor protein in renal cancer. Oncogene. 2010; 29:4914-4924.

59. Delforoush M, Strese S, Wickström M, Larsson R, Enblad G, Gullbo J. *In vitro* and *in vivo* activity of melflufen (J1) in lymphoma. BMC Cancer. 2016; 16:263.

60. Mcduffie HH, Pahwa P, Karunanayake CP, Spinelli JJ, Dosman JA. Clustering of cancer among families of cases with Hodgkin Lymphoma (HL), Multiple Myeloma (MM), Non-Hodgkin's Lymphoma (NHL), Soft Tissue Sarcoma (STS) and control subjects. BMC Cancer. 2009; 9:70.

61. Virginia Pascual FA, Pinakeen Patel, Karolina Palucka, Damien Chaussabel, Jacques Banchereau. How the Study of Children with Rheumatic Diseases Identified Interferon alpha and Interleukin 1 as Novel Therapeutic Targets. Immunol Rev. 2008; 223:39–59.

62. Chen S, Wang Z, Dai X, Pan J, Ge J, Han X, Wu Z, Zhou X, Zhao T. Re-expression of microRNA-150 induces EBV-positive Burkitt lymphoma differentiation by modulating c-Myb *in vitro*. Cancer Sci. 2013; 104:826–834.

63. Tian-Tian GE, Liang Y, Rong FU, Wang GJ, Er-Bao R, Wen QU, Wang XM, Liu H, Yu-Hong WU, Song J. Expressions of miR-21, miR-155 and miR-210 in Plasma of Patients withLymphoma and Its Clinical Significance. Zhongguo shi yan xue ye xue za zhi 2012; 20:305-309.

64. Gmyrek GA, Walburg M, Webb CP, Yu HM, You X, Vaughan ED, Vande Woude GF, Knudsen BS. Normal and malignant prostate epithelial cells differ in their response to hepatocyte growth factor/scatter factor. Am J Pathol. 2001; 159:579-590.

65. Walsh PC, Partin AW. Family history facilitates the early diagnosis of prostate carcinoma. Cancer. 1997; 80:1871-1874.

66. Hart M, Wach S, Nolte E, Szczyrba J, Menon R, Taubert H, Hartmann A, Stoehr R, Wieland W, Grässer FA. The proto-oncogene ERG is a target of microRNA miR-145 in prostate cancer. FEBS J. 2013; 280:2105–2116.

67. 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.

68. Bertsekas DP. (1999). Nonlinear programming: Athena scientific Belmont).

69. Candès EJ, Recht B. Exact Matrix Completion via Convex Optimization. Foundations of Computational Mathematics. 2008; 9:717-772.

70. Elman HC, Golub GH. Inexact and preconditioned Uzawa algorithms for saddle point problems. Siam J Numer Anal. 2010; 31:1645-1661.

71. Cai JF, Cand, S, Emmanuel J., Shen Z. A Singular Value Thresholding Algorithm for Matrix Completion. Siam Journal on Optimization. 2010; 20:1956-1982.

72. Kanehisa M, Goto S, Hattori M, Aokikinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M. From genomics to chemical genomics: new developments in KEGG. Nucleic Acids Res. 2006; 34:D354-D357.

73. Liao Q, Guan N, Wu C, Zhang Q. Predicting Unknown Interactions Between Known Drugs and Targets via Matrix Completion. Lecture Notes in Computer Science. 2016; 9651:591-604.