

In Silico Prediction of Small Molecule-miRNA Associations Based on the HeteSim Algorithm

Jia Qu,¹ Xing Chen,¹ Ya-Zhou Sun,² Yan Zhao,¹ Shu-Bin Cai,² Zhong Ming,² Zhu-Hong You,³ and Jian-Qiang Li²

¹School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China; ²College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China; ³Xinjiang Technical Institute of Physics and Chemistry, Chinese Academy of Science, Ürümqi 830011, China

Targeting microRNAs (miRNAs) with drug small molecules (SMs) is a new treatment method for many human complex diseases. Unsurprisingly, identification of potential miRNA-SM associations is helpful for pharmaceutical engineering and disease therapy in the field of medical research. In this paper, we developed a novel computational model of HeteSim-based inference for SM-miRNA Association prediction (HSSMMA) by implementing a path-based measurement method of HeteSim on a heterogeneous network combined with known miRNA-SM associations, integrated miRNA similarity, and integrated SM similarity. Through considering paths from an SM to a miRNA in the heterogeneous network, the model can capture the semantics information under each path and predict potential miRNA-SM associations based on all the considered paths. We performed global, miRNA-fixed local and SM-fixed local leave one out cross validation (LOOCV) as well as 5-fold cross validation based on the dataset of known miRNA-SM associations to evaluate the prediction performance of our approach. The results showed that HSSMMA gained the corresponding areas under the receiver operating characteristic (ROC) curve (AUCs) of 0.9913, 0.9902, 0.7989, and 0.9910 ± 0.0004 based on dataset 1 and AUCs of 0.7401, 0.8466, 0.6149, and 0.7451 ± 0.0054 based on dataset 2, respectively. In case studies, 2 of the top 10 and 13 of the top 50 predicted potential miRNA-SM associations were confirmed by published literature. We further implemented case studies to test whether HSSMMA was effective for new SMs without any known related miRNAs. The results from cross validation and case studies showed that HSSMMA could be a useful prediction tool for the identification of potential miRNA-SM associations.

INTRODUCTION

Functional studies revealed that RNAs, once thought to be simple messengers from DNA to protein, have important roles in many cellular processes.¹ They were found to regulate transcription, translation, RNA modification, mRNA stability, chromatin structure, and signaling pathways by interacting with various biological molecules.² Most of these processes are related to a variety of human diseases, including cancers and neurodegenerative and neuromuscular diseases.^{3,4} These discoveries have validated the potential of RNAs as therapeutic targets.⁵ Therefore, scientists have been excited about

the prospect of the new class of drug targets, and they have developed novel tools to study the drug ability of RNAs.

For decades, small molecules (SMs) have been studied for their modulatory function by therapeutic targeting of proteins.⁶ While there are many examples of protein-targeted drug design, similar research for RNA targeting wasn't uncovered until the late 1980s.⁷ Several classes of antibacterial drugs were found to be able to bind to bacterial rRNA to exert their function.⁸ To date, SMs have been found to target mammalian RNAs through various mechanisms. These RNAs are from various regions of the genome. They can be categorized into five general classes based on structure: RNA splicing, microRNAs (miRNAs), RNA repeat elements, G-quadruplex structures, and ribosomal synthesis.⁹

miRNAs are short 20- to 25-nt non-coding RNA (ncRNA) transcripts.¹⁰ They have important post-transcriptional cellular functions by binding to their target mRNA, resulting in mRNA decay or inhibition of translation.¹¹ Many miRNAs are correlated with a variety of human diseases, including diabetes; obesity; cancers; and neurodegenerative, autoimmune, and cardiovascular diseases.¹² Therefore, they are becoming potential SM drug targets.¹³ SMs that can bind to the Drosha- or Dicer-processing site target downstream miRNA to prevent the processing of mature miRNA formation or regulate function of miRNA. Several laboratories have reported SM inhibitors of Drosha- or Dicer-miRNA interactions. Disney et al.¹⁴ found that the molecule Targaprimir-96 could selectively target pri-miRNA 96 and, subsequently, inhibit Drosha processing in triple-negative breast cancer (TNBC). In a separate report, they found that ligand Targapremir-210 targeted pre-miRNA-210 and inhibited Dicer processing under hypoxic conditions.¹⁵ Another

Received 13 September 2018; accepted 4 December 2018;
<https://doi.org/10.1016/j.omtn.2018.12.002>.

Correspondence: Xing Chen, School of Information and Control Engineering, China University of Mining and Technology, No. 1, Daxue Road, Xuzhou 221116, China.

E-mail: xingchen@amss.ac.cn

Correspondence: Zhong Ming, College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China.

E-mail: mingz@szu.edu.cn

Correspondence: Zhu-Hong You, Xinjiang Technical Institute of Physics and Chemistry, Chinese Academy of Science, Ürümqi 830011, China.

E-mail: zhuhongyou@ms.xjb.ac.cn



search yielded a ligand bis-benzimidazole to target miRNA-544; by targeting a functional miRNA-processing site, it achieved successful miRNA-544 inhibition.¹⁶

To identify SM-RNA associations, several high-throughput-screening approaches have been developed. They are based on mass spectrometry, fluorescence, or reporter.^{17–20} Currently, reporter-based assays are most commonly used in the identification of miRNA-targeted SM inhibitors.¹² The principle of this method is that the luciferase activity can be suppressed when miRNA binds to the 3' UTR sequence of its target luciferase gene. If SM can target miRNA and decrease the binding, the luciferase readout of the labeled mRNA will increase.¹² However, this assay could not provide information about specific miRNA-gene interaction or direct miRNA engagement that the SM may inhibit.

On the other hand, several informatics-based high-throughput-screening methods that measure direct miRNA binding have been developed to aid miRNA-targeted drug discovery. For example, through two-dimensional combinatorial screening (2DCS)²¹ and statistical method analyzing structure-activity relationships based on sequence (StARTS),²² Disney et al.¹⁴ constructed the RNA motif-SM database named Inforna. These approaches have aided in the discovery of miRNA-SM association, though the further mechanisms of association were revealed during follow-up biological experiments rather than during the screening. Recently, Lorenz and Garner²³ reported a novel high-throughput-screening approach that may compensate for these drawbacks. This approach can detect direct inhibitors of miRNA cleavage, and it may thus be promising for discovering SMs that target specific pre-miRNA.

The research of miRNA-SM associations has significantly advanced over the past few years. Continued development of both fundamental mechanisms and biological approaches has further benefited the discovery of miRNA-targeted SM ligands. However, the identification and validation of miRNA-SM associations by biological experiment are still always time consuming and costly. It is urgent for the scientists to elucidate how to select the best-suited miRNAs for targeting by more efficient and accurate methods. Developing computational prediction models for miRNA-SM associations is a promising strategy. The efficient and reliable prediction models can lead to the efficient design of miRNA-targeted ligands. Predicting miRNA-SM association will also be crucial for the selectivity of SM targeting miRNA and establishing guidelines for SM targeting miRNA research. This work would significantly aid in our understanding of the relationships between SMs and miRNAs. Also, on the basis of the association information between miRNAs and SMs, we could further predict synergistic drug combinations.²⁴ Therefore, there is an urgent need to develop new computational approaches or models to speed up the studies of this field.

In recent years, a number of studies based on computational models have been developed on the identification of potential miRNA-SM associations, which may be helpful to a new direction of miRNA-tar-

geted therapies.^{12,25–27} Wang et al.²⁸ introduced a new method to predict potential miRNA-SM associations through calculating a functional similarity score of each miRNA-SM pair. In the model, they identified differentially expressed genes for drug treatment and miRNA perturbation, and then they calculated the functional similarity of each miRNA-SM pair by using gene ontology (GO) enrichment analysis on their differentially expressed genes. Meng et al.²⁹ proposed a computational approach to identify the potential associations between SMs and Alzheimer's disease (AD)-related miRNAs based on the differentially expressed target genes that were regulated by the aberrantly expressed AD-related miRNAs. Differentially expressed target genes used in the model were divided into two groups based on whether they were overexpressed or under-expressed, and Kolmogorov-Smirnov (KS) values for the overexpressed and under-expressed target genes were calculated respectively to obtain final scores for the potential associations between SMs and AD-related miRNAs by integrating the two KS values. Besides, Jiang et al.³⁰ have done the similar work for the identification of potential miRNA-SM associations in 23 different cancers through implementing a KS test based on the differential expression of miRNA target genes and transcriptional responses following drug treatment. They also constructed the SM-miRNA network for each cancer to analyze the property of each association, which could be helpful to the identification of drug candidates for cancer therapy.

Lv et al.³¹ developed a network-based computational method to predict the potential associations between miRNAs and SMs by implementing random walk with restart algorithm (RWR) on the integrated network that is composed of known miRNA-SM associations and SM-SM and miRNA-miRNA similarity. In addition, Li et al.³² also proposed a network-based framework named predictive SM-miRNA Network-Based Inference (SMiR-NBI), which established a network by connecting drugs, miRNAs, and genes and implemented a network-based inference (NBI) framework on the network to prioritize miRNAs for SMs. Recently, Qu et al.³³ put forward a calculation approach of Triple Layer Heterogeneous Network based SM-miRNA Association prediction (TLHNSMMA), in which they collected heterogeneous data about SMs, miRNAs, and diseases and treated disease information as a bridge to build a three-layer mixing network. On this basis, two iterative updating processes that spread information of heterogeneous data were generated to infer novel miRNA-SM associations and miRNA-disease associations simultaneously.

In this paper, drawing on previous research on the development of computational models on the miRNA-disease association prediction,^{34–38} we introduced a novel computational model of HeteSim-based inference for SM-miRNA Association prediction (HSSMMA) to calculate the relevance between miRNAs and SMs by implementing a path-based measurement method of HeteSim in a heterogeneous network that was constructed based on integrated miRNA similarity, integrated SM similarity, and experimentally confirmed miRNA-SM associations. The HeteSim is a path-constrained measurement method to calculate the relatedness of objects with the same or different types in a heterogeneous network based on the

search path that connect two objects by a sequence of node types.³⁹ In addition, the dataset of experimentally confirmed miRNA-SM associations used in this model was downloaded from the database of SM2miR.⁴⁰ According to the known miRNA-SM associations, we constructed two types of datasets (dataset 1 and dataset 2), and we implemented HSSMMA on them for the identification of potential miRNA-SM associations, respectively. In dataset 1, only a part of SMs and miRNAs are involved in the known miRNA-SM associations. In dataset 2, all SMs and miRNAs are involved in the known miRNA-SM associations.

To test the prediction performance of HSSMMA, we employed global, miRNA-fixed local and SM-fixed local leave one out cross validation (LOOCV) as well as 5-fold cross validation on the two datasets, respectively. The results showed that the area under the receiver operating characteristic (ROC) curve (AUC) of global LOOCV was 0.9913 and 0.7401 based on dataset 1 and dataset 2, respectively. Through fixing each miRNA to predict miRNA-associated SMs, the AUC of local LOOCV was, respectively, 0.9902 and 0.8466 for the dataset 1 and dataset 2. Through fixing each SM to predict SM-associated miRNAs, the AUC of local LOOCV was, respectively, 0.7989 and 0.6149 for the dataset 1 and dataset 2. For 5-fold cross validation, the average AUC and corresponding SD was 0.9910 ± 0.0004 and 0.7451 ± 0.0054 for the dataset 1 and dataset 2, respectively. In case studies, the results showed that 2 of the top 10 and 13 of the top 50 predicted miRNA-SM associations were confirmed by published references. We further implemented our model to the SMs without any known associated miRNAs. The known related miRNAs for the investigated SM would be removed from the training dataset and 5-fluorouracil, 17 β -estradiol, and 5-aza-2'-deoxycytidine were taken as the investigated SMs, respectively. We found that 27, 24, and 26 of the top 50 predicted miRNAs for 5-fluorouracil (5-FU), 17 β -estradiol (E2), and 5-aza-2'-deoxycytidine (5-Aza-CdR) were confirmed by SM2miR database and published literature reports, respectively.

RESULTS

Performance Evaluation

For evaluating the prediction performance of HSSMMA, global, miRNA-fixed local, and SM-fixed local LOOCV as well as 5-fold cross validation were implemented based on the dataset of known miRNA-SM associations. We further compared the performance of HSSMMA with one previous classical computational model of SMiR-NBI based on dataset 1 and dataset 2, respectively. In LOOCV, we regarded each known miRNA-SM pair as a test sample in turn; the remaining known associations between miRNAs and SMs were considered as training samples. The miRNA-SM pairs without known associations were used as candidate samples. We would obtain the prediction scores of each miRNA-SM pair after HSSMMA was implemented. In the global LOOCV evaluation, we would compare the score of the test sample with the scores of all the candidate samples. In the miRNA-fixed local LOOCV, we would sort the score of the test sample with the scores of candidate samples that were made up of the pairs between SMs and fixed miRNAs. Meanwhile, in the SM-fixed local LOOCV, we would sort the score of the test sample with the

scores of candidate samples that were made up of the all pairs between miRNAs and fixed SMs.

In 5-fold cross validation, we would randomly divide the known miRNA-SM associations into five equal parts; one part was selected as the test sample in turn and the rest (four parts) were regarded as training samples. In the same way, the miRNA-SM pairs without any known associations would be considered as candidate samples and the score of each test sample would be ranked with all candidate samples, respectively. Because of the random partition of the original samples in 5-fold cross validation, the corresponding validation results would be different in each random process. Therefore, we repeated the process of 5-fold cross validation 100 times in this study, and we took the average of the 100 cross validation results to evaluate the model.

At last, we plotted the ROC curve using true positive rate (TPR, sensitivity) against the false positive rate (FPR, 1-specificity) at different thresholds. Sensitivity denotes the percentage of test samples that were identified with higher ranks than the given threshold. Specificity refers to the percentage of negative miRNA-SM pairs with lower ranks than the threshold. After that, we calculated AUC as an evaluation index for the prediction performance of HSSMMA. If the value of AUC is 0.5, the prediction performance of the HSSMMA is random; if the value of AUC is 1, the prediction performance of HSSMMA is perfect. In the global LOOCV, the results showed that HSSMMA and SMiR-NBI obtained AUCs of 0.9913 and 0.8843 based on dataset 1, respectively, and AUCs of 0.7401 and 0.7264 based on dataset 2, respectively (see [Figure 1](#)). In the framework of miRNA-fixed local LOOCV, HSSMMA and SMiR-NBI obtained AUCs of 0.9902 and 0.8837, respectively, based on dataset 1 and AUCs of 0.8466 and 0.7846, respectively, based on dataset 2 (see [Figure 2](#)). In the framework of SM-fixed local LOOCV, HSSMMA and SMiR-NBI obtained AUCs of 0.7989 and 0.7497, respectively, based on dataset 1 and AUCs of 0.6149 and 0.6100, respectively, based on dataset 2 (see [Figure 3](#)). In 5-fold cross validation, HSSMMA and SMiR-NBI obtained AUCs of 0.9910 ± 0.0004 and 0.8554 ± 0.0063 , respectively, based on dataset 1 and AUCs of 0.7451 ± 0.0054 and 0.7104 ± 0.0087 , respectively, based on dataset 2. Comparisons between evaluation results of HSSMMA and SMiR-NBI demonstrate that HSSMMA is reliable and effective for the identification of potential miRNA-SM associations.

Case Studies

We carried out a case study based on the dataset 1 to evaluate the effectiveness of the HSSMMA. After the implementation of HSSMMA, we observed the number of the verified miRNA-SM associations in the top 10, top 20, and top 50 prediction list. As the results showed, among the top 10, 20, and 50 potential SM-miRNA associations, there were 2, 4, and 13 associations confirmed by experiments, respectively (see [Table 1](#)). It is worth noting that the SM was presented using PubChem compound identifier (PubChem-CID) in the dataset of known miRNA-SM associations.

For instance, in the top 10 predicted miRNA-SM associations, the predicted association between mir-203a and 5-Aza-CdR is ranked

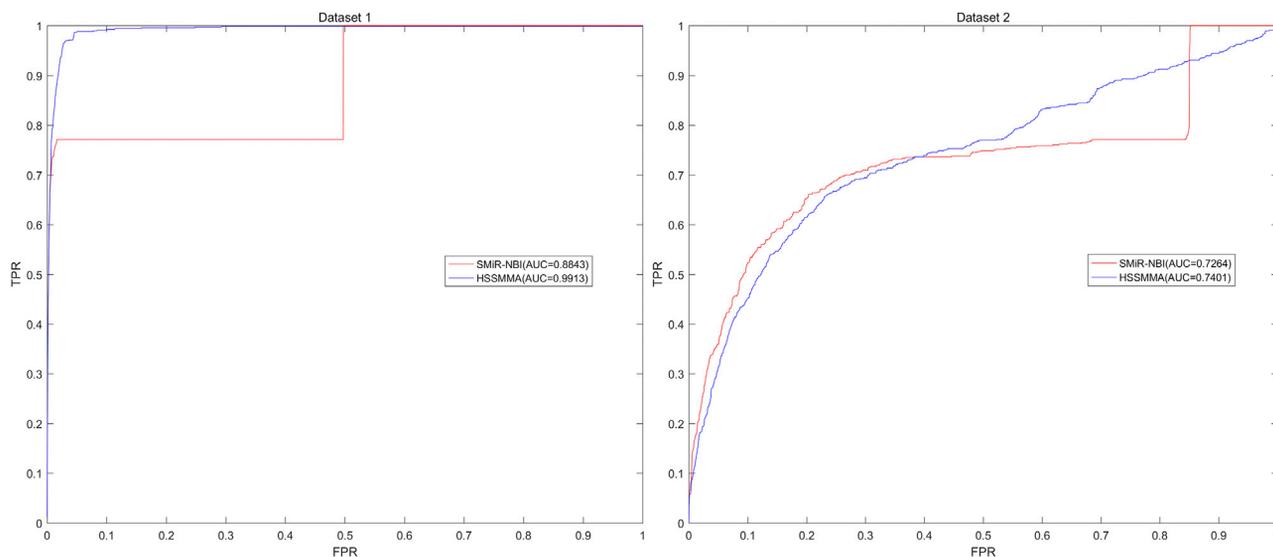


Figure 1. Performance Comparison between HSSMMA and SMIR-NBI in Terms of ROC Curves and AUCs Based on Global LOOCV

HSSMMA obtained AUCs of 0.9913 and 0.7401 based on dataset 1 and dataset 2, respectively.

first. Recent study revealed that treatment of esophageal squamous cell carcinoma (ESCC) cells with 5-Aza-CdR resulted in increased miR-203a expression.⁴¹ The seventh predicted association is between let-7d and 5-Aza-CdR. Zhang et al.⁴² found that let-7d could inhibit dopamine D3 receptor (DRD3) expression in immortalized renal proximal tubule (RPT) cells via methylation. This inhibition could be abolished by 5-Aza-CdR. In the top 20 predicted miRNA-SM associations, we also revealed the potential association between miR-128-1 and 5-Aza-CdR ranked 15th. Shan et al.⁴³ found that miR128-1 is downregulated and closely associated with glioblastoma multiforme (GBM). Treatment with the 5-Aza-CdR resulted in miR128-1 upregulation in GBM cells, and it inhibited tumor cell proliferation, suggesting that 5-Aza-CdR may potentially treat GBM by upregulating miR-128-1.⁴³ Besides, the 20th predicted association between let-7a-3 and 5-Aza-CdR was verified. Zhu et al.⁴⁴ found that treatment with 5-Aza-CdR could decrease the methylation density of let-7a-3 promoter and increase the level of let-7a-3 expression in acute myeloid leukemia (AML) cells. We further used HSSMMA to predict candidate miRNAs for all SMs in dataset 1, and the corresponding probability scores are provided (see Table S1).

To validate the prediction effectiveness for new SMs without any known related miRNAs, we carried out another case study based on dataset 1 by removing all associations between miRNAs and the investigated SM and implementing HSSMMA based on the rest of the known miRNA-SM associations. Then all predicted miRNAs for the investigated SM would be ranked according to their predicted scores, and the top 50 potential associations between miRNAs and the investigated SM would be confirmed by the SM2miR database and published references. Here, 5-FU, E2, and 5-Aza-CdR were taken as the investigated SMs, respectively.

5-FU

The agent 5-FU is a widely used chemotherapeutic drug in cancers.⁴⁵ It induces cytotoxic effects by altering DNA and RNA metabolism and mRNA expression.^{46–48} Exposure to 5-FU promotes transcriptional reprogramming, leading to the alteration of mRNA or miRNA expression profiles that results in a change of cell fate.^{49–51} After implementing HSSMMA, we got the total ranking of potential miRNAs associated with 5-FU. As the results show, among the top 10 potential 5-FU-related miRNAs, there are 7 confirmed miRNAs, among which 6 miRNAs were confirmed by the known SM-miRNA association dataset constructed in the literature³¹ and 1 miRNA was confirmed by experimental report from the literature.⁵² Among the top 50 potential 5-FU-related miRNAs, there are 27 confirmed miRNAs, among which 19 miRNAs were confirmed by the known SM-miRNA association dataset constructed in the literature³¹ and 8 miRNAs were confirmed by experimental report from the literature (see Table 2).

For instance, Huang et al.⁵² found that knockdown of astrocyte-elevated gene-1 (AEG-1) in colorectal cancer (CRC) cells can improve the expression of miR-181a-2 and increase the sensitivity of CRC cells to 5-FU, suggesting a potential mechanism to improve the efficiency of 5-FU by miR-181a-2. The association between let-7b and 5-FU was confirmed by Wang et al.⁵³ They found that let-7b can decrease the expression of B cell lymphoma-extra large (Bcl-xL) and sensitize breast cancer cells to 5-FU.⁵³ Through functional assays, Han et al.⁵⁴ revealed that restoration of miR-874 could inhibit the proliferation of the CRC cells and decrease the 5-FU resistance of the CRC cells. In TNBC, the expression level of miR-205 was significantly elevated in MDA-MB-453 LAR-type TNBC tumor cells treated with 5-FU together with ixabepilone, suggesting the drugs may exert effects through the regulation of miR-205.⁵⁵

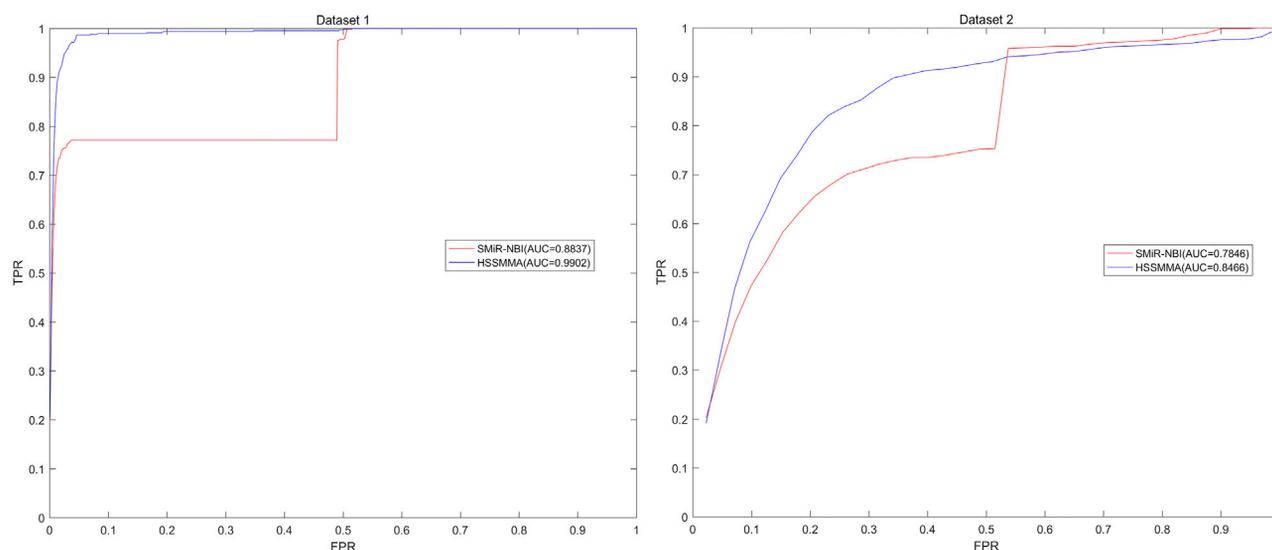


Figure 2. Performance Comparison between HSSMMA and SMiR-NBI in Terms of ROC Curves and AUCs Based on Local LOOCV by Fixing miRNAs to Rank SMs

HSSMMA obtained AUCs of 0.9902 and 0.8466 based on dataset 1 and dataset 2, respectively.

E2

The small molecular E2 is a type of human estrogen that exerts important effects on the reproductive and other related biological processes in both men and women.⁵⁶ It has significant anticancer activity against human breast cancer.⁵⁷ After implementing HSSMMA, we got the total ranking of miRNAs for E2. In the top 10 miRNAs, 5 miRNAs were confirmed by the database of SM2miR³¹ and 1 miRNA was confirmed by previous literature.⁵⁸ Moreover, 14 of the top 50 miRNAs were verified by SM2miR and 10 were confirmed by experimental literature (see Table 3). For example, Aqil et al.⁵⁸ confirmed that the overexpression of miR-375 following E2 treatment could be significantly protected by jamun diet. Li et al.⁵⁹ demonstrated that miR-22 could effectively reverse E2-induced cell proliferation and invasion of tumor cells in endometrial endometrioid carcinomas (EECs) by inhibiting Cyclin D1 expression and the secretion of matrix metalloproteinase (MMP)-2 and MMP-9. Zhang et al.⁶⁰ reported that miR-320a expression is significantly downregulated in cumulus granulosa cells (CCs) from polycystic ovary syndrome (PCOS) patients and this downregulation promotes E2 deficiency in CCs. Treatment of E2-sensitive MCF7 breast cancer cells with fulvestrant resulted in increased expression of endogenous miR-221.⁶¹

5-Aza-CdR

The ligand 5-Aza-CdR is an inhibitor of DNA methyltransferase (DNMT). It can reverse methylation and reactivate the expression of silenced genes.⁶² 5-Aza-CdR was able to suppress the growth of various tumors.^{63–65} We performed HSSMMA on 5-Aza-CdR and got the total ranking of all 5-Aza-CdR-associated miRNAs. The results showed that 5 of the first 10 miRNAs were confirmed by the database of SM2miR³¹ and 1 was confirmed by previous literature.⁴¹

Furthermore, 17 of the top 50 miRNAs were verified by SM2miR and 9 were confirmed by experiments (see Table 4). For instance, Xu et al.⁶⁶ found that the expression level of let-7b was significantly reduced in acute lymphoblastic leukemia (ALL) patients. The 5-Aza-CdR could increase the expression of let-7b and inhibit the growth of ALL cells.⁶⁶ Through bisulfite pyrosequencing of bladder cancer (BCa) cell lines treated with 5-Aza-CdR and 4-phenylbutyric acid (PBA), Shimizu et al.⁶⁷ identified upregulated miRNAs by 5-aza-dC plus PBA. Among them, miR-124-2 and miR-124-3 were frequently and tumor-specifically methylated in primary cancers.⁶⁷ In breast cancer, Manavalan et al.⁶⁸ found that 5-Aza-CdR in combination with histone deacetylase inhibitor trichostatin A (TSA) could increase miR-200c in LY2 cells. In GBM, Ghasemi et al.⁶⁹ revealed that treatment of U87MG cells with 5-Aza-CdR can reverse the hypermethylation status of miR-149 and increase its expression, thus decreasing target mRNA and protein levels.

DISCUSSION

miRNAs as a new potential therapeutic target have attracted wide attention.⁷⁰ Moreover, large quantities of studies have certified that SMs could modulate the expression of miRNAs and, thus, have potential for treating diseases.²⁸ Not surprisingly, the identification of potential miRNA-SM associations has important sense for disease therapy and drug clinical applications. In this study, we introduced a computational model of HSSMMA to predict potential miRNA-SM associations by implementing a path-constrained measurement method of HeteSim on a heterogeneous network that was established with known miRNA-SM associations, miRNA-miRNA similarity, and SM-SM similarity. The results of cross validation and case studies showed that the model could effectively predict potential miRNA-SM associations.

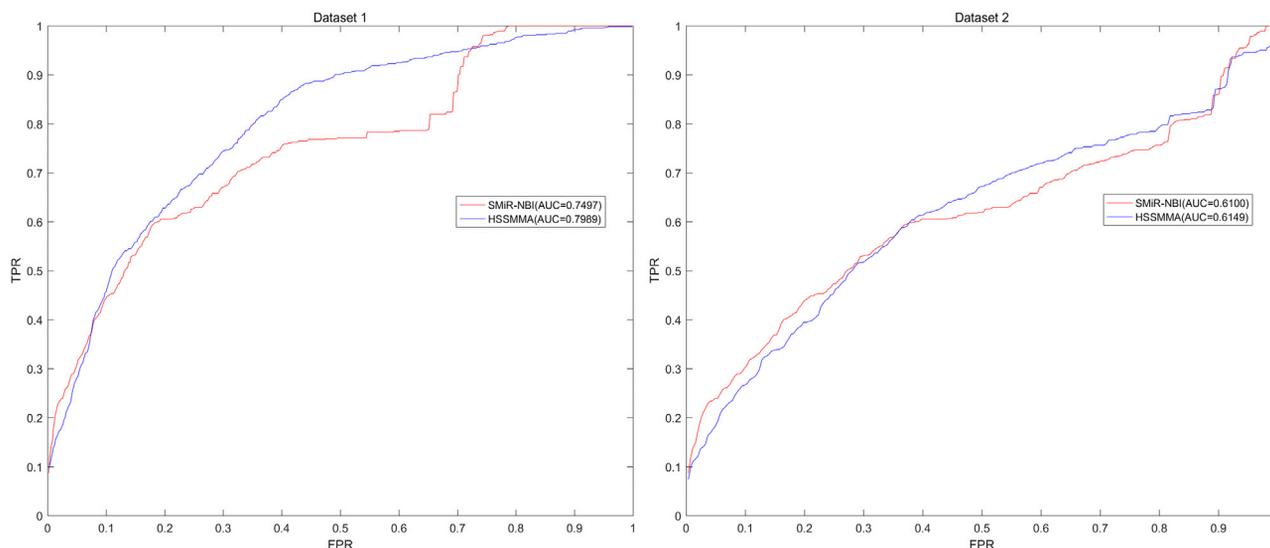


Figure 3. Performance Comparison between HSSMMA and SMIR-NBI in Terms of ROC Curves and AUCs Based on Local LOOCV by Fixing SMs to Rank miRNAs

HSSMMA obtained AUCs of 0.7989 and 0.6149 based on dataset 1 and dataset 2, respectively.

The effectiveness of HSSMMA is probably traceable in part to the following several factors. First, the dataset of known miRNA-SM associations used in this paper was collected from a highly reliable SM2miR database. Several other reliable datasets were also used in the model, such as SM side effect similarity, SM chemical structure similarity, gene functional consistency-based similarity for SMs and miRNAs, and disease phenotype-based similarity for miRNAs and SMs, which can greatly increase algorithm efficiency. Second, HSSMMA is a path-based relevance measure, which can effectively capture the subtle semantics information of search paths. Therefore, through selecting search paths from an SM to a miRNA in the established heterogeneous network, HSSMMA can be a useful prediction tool for prioritizing potential miRNA-SM associations. More importantly, compared with some machine-learning-based models that randomly select negative samples as training data, HSSMMA only requires positive samples as training data. The random selection of negative samples in machine-learning-based models would affect their prediction accuracy. Consequently, the prediction accuracy of the HSSMMA model is more convincing than that of the prediction models with negative samples as training data.

However, there are also some weaknesses in our proposed model: the 664 experimentally confirmed miRNA-SM associations used in this paper are far from enough. More known miRNA-SM associations need to be verified by experiment, which would contribute to the improvement of prediction accuracy for the model. Furthermore, the function for the combination of several different HeteSim scores based on search paths in our approach is relatively simple, and it could be reconstructed based on advanced machine-learning methods. Compared with previous models, HSSMMA significantly improved prediction ability for the identification of potential miRNA-SM asso-

ciations. However, the prediction accuracy of HSSMMA is still not satisfactory. These factors all motivate researchers to develop more effective computational models to predict potential miRNA-SM associations based on the reliable biological datasets.

MATERIALS AND METHODS

SM-miRNA Associations

The dataset of 664 experimentally verified miRNA-SM associations used in this study was downloaded from the SM2miR database.⁴⁰ Dataset 1 includes 541 miRNAs, 831 SMs, and 664 known miRNA-SM associations, of which only a portion of miRNAs and SMs participate in the known miRNA-SM associations. Dataset 2 includes 39 SMs, 286 miRNAs, and 664 known miRNA-SM associations, where all SMs and miRNAs participate in the known associations. All the known SM-miRNA associations are listed in Table S2. More importantly, we constructed an adjacency matrix A to indicate the known miRNA-SM associations. If SM i is associated with miRNA j , the entity $A(i,j)$ is 1, otherwise 0. We further used ns and nm to indicate the number of SMs and miRNAs in the dataset.

Integrated SM Similarity

Lv et al.³¹ calculated an SM similarity score based on SM side effect similarity S_S^s ,⁷¹ gene functional consistency-based similarity for SMs S_S^T ,⁷² SM chemical structure similarity S_S^C ,⁷³ and disease phenotype-based similarity for SMs S_S^D .⁷¹ Here, SM side effect similarity was computed by using the Jaccard score⁷¹ based on the dataset of SM drug side effects that was collected from SDe Effect Resource (SIDER).⁷⁴ Gene functional consistency-based similarity for SMs was reflected by employing the Gene Set Functional Similarity (GSFS) method⁷² on the dataset of target genes of the SMs obtained from DrugBank⁷⁵ and Therapeutic Targets Database (TTD).⁷⁶

Table 1. Prediction of the Top 50 Potential miRNA-Related SMs Based on Dataset 1

SM	miRNA	Evidence	SM	miRNA	Evidence
CID 451668	hsa-mir-203a	26577858	CID 9444	hsa-mir-21	unconfirmed
CID 451668	hsa-mir-23a	unconfirmed	CID 451668	hsa-let-7b	26708866
CID 451668	hsa-mir-15b	unconfirmed	CID 451668	hsa-let-7e	22053057
CID 451668	hsa-let-7a-1	unconfirmed	CID 451668	hsa-mir-199a-1	unconfirmed
CID 60750	hsa-mir-23a	unconfirmed	CID 451668	hsa-mir-152	unconfirmed
CID 9444	hsa-mir-106b	unconfirmed	CID 3385	hsa-mir-181a-1	unconfirmed
CID 451668	hsa-let-7d	26802971	CID 451668	hsa-mir-146a	24885368
CID 451668	hsa-mir-132	unconfirmed	CID 3385	hsa-let-7b	25789066
CID 9444	hsa-mir-93	unconfirmed	CID 36462	hsa-let-7f-1	unconfirmed
CID 9444	hsa-mir-16-1	unconfirmed	CID 5288826	hsa-mir-17	unconfirmed
CID 451668	hsa-mir-128-2	unconfirmed	CID 9444	hsa-let-7f-1	unconfirmed
CID 9444	hsa-mir-16-2	unconfirmed	CID 3229	hsa-let-7g	unconfirmed
CID 451668	hsa-mir-18a	unconfirmed	CID 60750	hsa-mir-17	23001407
CID 9444	hsa-mir-210	unconfirmed	CID 3229	hsa-let-7e	unconfirmed
CID 451668	hsa-mir-128-1	27705931	CID 9444	hsa-mir-191	unconfirmed
CID 9444	hsa-mir-15a	unconfirmed	CID 60750	hsa-mir-24-2	25841339
CID 451668	hsa-let-7a-2	unconfirmed	CID 451668	hsa-mir-30e	unconfirmed
CID 451668	hsa-mir-199a-2	unconfirmed	CID 451668	hsa-mir-26b	unconfirmed
CID 451668	hsa-mir-92a-1	unconfirmed	CID 3385	hsa-mir-26a-1	unconfirmed
CID 451668	hsa-let-7a-3	26227220	CID 3385	hsa-mir-155	28347920
CID 9444	hsa-mir-25	unconfirmed	CID 451668	hsa-mir-342	unconfirmed
CID 451668	hsa-mir-106a	unconfirmed	CID 5288826	hsa-let-7f-1	unconfirmed
CID 3229	hsa-mir-24-1	unconfirmed	CID 9444	hsa-mir-23b	unconfirmed
CID 451668	hsa-mir-26a-1	unconfirmed	CID 451668	hsa-mir-200b	23626803
CID 3385	hsa-mir-126	26062749	CID 36462	hsa-mir-23b	unconfirmed

The SM was presented using the PubChem compound identifier (CID) in the dataset of known miRNA-SM associations. The first two columns record the top 1–25 miRNA-SM associations. The fourth and fifth columns record the top 26–50 miRNA-SM associations. The evidence for the associations was recent experimental literature with the corresponding PubMed ID.

Moreover, SM chemical structure similarity was gained by enforcing a graph-based method of SIMilar COMpound (SIMCOMP),⁷³ in light of the dataset of SM chemical structure extracted from the drug and compound sections of the Kyoto Encyclopedia of Genes and Genomes (KEGG) ligand database.⁷⁷ At last, based on the dataset of SM-related diseases downloaded from DrugBank and TTD, disease phenotype-based similarity for SMs was defined by employing Jaccard score.

To balance the four types of SM similarity and reduce the deviation of each similarity, the integrated SM similarity S_S can be defined as follows:

$$S_S = (\beta_1 S_S^D + \beta_2 S_S^T + \beta_3 S_S^C + \beta_4 S_S^S) / \sum_j \beta_j (j = 1, 2, 3, 4), \quad (\text{Equation 1})$$

where the default value $\beta_j = 1$ means that the separated similarities have the same weight. In this paper, we have released the integrated similarity of all 831 SMs that were used in dataset 1 (see Table S3).

Integrated miRNA Similarity

In the same way, Lv et al.³¹ also calculated miRNA similarity by a weighed combination of gene functional consistency-based similarity for miRNAs S_M^T and disease phenotype-based similarity for miRNAs S_M^D .^{71,72} On the basis of the target genes of each miRNA from TargetScan,⁷⁸ gene functional consistency-based similarity for miRNAs could be gained through carrying out the GFSF method⁷² to compute functional consistency of their miRNA target gene sets;⁷² disease phenotype-based similarity for miRNAs was calculated by using Jaccard score according to the dataset of miRNA-related diseases extracted from Human MicroRNA Disease Database (HMDD) version (v.)2.0⁷⁹ database and miR2Disease⁸⁰ and PhenomiR⁸¹ databases. At last, the integrated miRNA similarity S_M can be defined as follows:

$$S_M = (\alpha_1 S_M^D + \alpha_2 S_M^T) / \sum_i \alpha_i (i = 1, 2), \quad (\text{Equation 2})$$

where the default value $\alpha_i = 1$ means the separated similarities possess the same weight. Likewise, we have released the integrated similarity of all 541 miRNAs that were used in dataset 1 (see Table S4).

Table 2. Top 50 miRNAs Associated with 5-Fluorouracil Were Predicted by HSSMMA Based on Dataset 1

SM	miRNA	Evidence	SM	miRNA	Evidence
CID 3385	hsa-mir-324	unconfirmed	CID 3385	hsa-mir-155	28347920
CID 3385	hsa-mir-24-1	26198104	CID 3385	hsa-mir-320a	26198104
CID 3385	hsa-mir-23a	26198104	CID 3385	hsa-mir-126	26062749
CID 3385	hsa-mir-24-2	26198104	CID 3385	hsa-mir-1226	26198104
CID 3385	hsa-mir-500a	unconfirmed	CID 3385	hsa-mir-409	unconfirmed
CID 3385	hsa-mir-501	26198104	CID 3385	hsa-mir-197	26198104
CID 3385	hsa-mir-181a-1	unconfirmed	CID 3385	hsa-mir-27b	26198104
CID 3385	hsa-mir-21	26198104	CID 3385	hsa-mir-345	unconfirmed
CID 3385	hsa-mir-27a	26198104	CID 3385	hsa-mir-132	26198104
CID 3385	hsa-mir-181a-2	24462870	CID 3385	hsa-let-7d	26198104
CID 3385	hsa-let-7b	25789066	CID 3385	hsa-mir-199a-2	26198104
CID 3385	hsa-mir-874	27221209	CID 3385	hsa-mir-128-2	26198104
CID 3385	hsa-mir-16-1	26198104	CID 3385	hsa-mir-299	unconfirmed
CID 3385	hsa-let-7a-1	26198104	CID 3385	hsa-mir-205	24396484
CID 3385	hsa-mir-650	unconfirmed	CID 3385	hsa-mir-373	-
CID 3385	hsa-mir-125b-1	unconfirmed	CID 3385	hsa-mir-128-1	26198104
CID 3385	hsa-mir-26a-1	unconfirmed	CID 3385	hsa-mir-342	26198104
CID 3385	hsa-mir-125b-2	unconfirmed	CID 3385	hsa-mir-194-1	unconfirmed
CID 3385	hsa-mir-124-1	unconfirmed	CID 3385	hsa-let-7c	25951903
CID 3385	hsa-mir-181b-1	unconfirmed	CID 3385	hsa-mir-149	26198104
CID 3385	hsa-mir-328	unconfirmed	CID 3385	hsa-mir-186	unconfirmed
CID 3385	hsa-mir-124-2	unconfirmed	CID 3385	hsa-mir-154	unconfirmed
CID 3385	hsa-mir-124-3	unconfirmed	CID 3385	hsa-mir-204	27095441
CID 3385	hsa-mir-346	unconfirmed	CID 3385	hsa-mir-337	unconfirmed
CID 3385	hsa-mir-181b-2	unconfirmed	CID 3385	hsa-mir-1-2	28347920

The top 1–25 miRNAs are shown in the second column while the top 26–50 are in the fifth column. As a result, 7 and 27 of the top 10 and top 50 were confirmed by the known experimental literature, respectively.

HSSMMA

In this study, motivated by previous research introduced by Zheng et al.,⁸² we combined integrated SM similarity, integrated miRNA similarity, and known miRNA-SM associations into a heterogeneous network (see Figure 4 step 1), and we proposed a computational method of HSSMMA to infer potential associations between miRNAs and SMs by implementing HeteSim on an established heterogeneous network. Here, HeteSim is a path-based measurement method that evaluates the relatedness of object pairs depending on the considered search paths that connect two objects through a sequence of node types.³⁹ Unlike homogeneous networks, the paths in heterogeneous networks possess semantics. Consequently, on the basis of a basic idea that similar objects are relevant to similar objects, Kong and colleagues³⁹ designed the uniform and symmetric measure of HeteSim for arbitrary paths to compute the associations of heterogeneous objects. For a given path (symmetric or asymmetric), the measure can calculate the association score of each heterogeneous object pair. Moreover, HeteSim has been implemented in the identification of lncRNA-protein interactions,⁸² dis-

ease-gene associations,⁸³ and lncRNA-protein associations,⁸⁴ which demonstrated the high precision and good performance of HeteSim. Therefore, we carried out HeteSim measurement to infer potential miRNA-SM associations.

SM and miRNA are two types of objects in the miRNA-SM heterogeneous network. $SM \xrightarrow{R} miRNA$ indicates a relationship from type SM to type miRNA and the length l of the path is 1, where SM and miRNA represent the source type and target type of relationship R , respectively. A is an adjacent matrix between SM and miRNA, and the row-normalized matrix of A can be defined as follows:

$$T_{sm}(i, j) = \frac{A(i, j)}{\sum_{k=1}^{nm} A(i, k)}, \quad (\text{Equation 3})$$

where T_{sm} can be regarded as the transition probability matrix from SM to miRNA. In the same way, T_{ss} and T_{mm} can be calculated to indicate the transition probability matrix from SM to SM and miRNA to miRNA by obtaining the row-normalized matrix of integrated SM similarity S_S and integrated miRNA similarity S_M , respectively.

Table 3. Top 50 miRNAs Associated with 17 β -Estradiol Were Predicted by HSSMMA Based on Dataset 1

SM	miRNA	Evidence	SM	miRNA	Evidence
CID 5757	hsa-mir-375	27030099	CID 5757	hsa-mir-146b	29331043
CID 5757	hsa-mir-324	unconfirmed	CID 5757	hsa-mir-17	26198104
CID 5757	hsa-mir-23a	26198104	CID 5757	hsa-mir-125b-1	unconfirmed
CID 5757	hsa-mir-21	26198104	CID 5757	hsa-mir-345	unconfirmed
CID 5757	hsa-mir-194-1	unconfirmed	CID 5757	hsa-mir-125b-2	unconfirmed
CID 5757	hsa-mir-24-1	unconfirmed	CID 5757	hsa-mir-195	unconfirmed
CID 5757	hsa-mir-124-1	26198104	CID 5757	hsa-mir-203a	26198104
CID 5757	hsa-mir-181a-1	unconfirmed	CID 5757	hsa-mir-34a	24050776
CID 5757	hsa-mir-124-2	26198104	CID 5757	hsa-mir-20a	21914226
CID 5757	hsa-mir-124-3	26198104	CID 5757	hsa-mir-335	unconfirmed
CID 5757	hsa-mir-346	unconfirmed	CID 5757	hsa-mir-196a-1	unconfirmed
CID 5757	hsa-mir-22	24715036	CID 5757	hsa-mir-663a	26198104
CID 5757	hsa-mir-194-2	unconfirmed	CID 5757	hsa-mir-130b	unconfirmed
CID 5757	hsa-mir-16-1	unconfirmed	CID 5757	hsa-mir-92a-1	unconfirmed
CID 5757	hsa-mir-27b	26198104	CID 5757	hsa-mir-370	unconfirmed
CID 5757	hsa-mir-27a	26198104	CID 5757	hsa-mir-373	unconfirmed
CID 5757	hsa-mir-320a	27965096	CID 5757	hsa-mir-25	unconfirmed
CID 5757	hsa-mir-26a-1	unconfirmed	CID 5757	hsa-mir-106b	28422740
CID 5757	hsa-mir-15b	26198104	CID 5757	hsa-mir-152	unconfirmed
CID 5757	hsa-mir-221	21057537	CID 5757	hsa-mir-15a	unconfirmed
CID 5757	hsa-mir-126	26198104	CID 5757	hsa-mir-18a	24245576
CID 5757	hsa-mir-16-2	-	CID 5757	hsa-mir-150	-
CID 5757	hsa-mir-29a	22334722	CID 5757	hsa-mir-9-1	26198104
CID 5757	hsa-mir-26a-2	unconfirmed	CID 5757	hsa-mir-9-2	26198104
CID 5757	hsa-mir-24-2	27030099	CID 5757	hsa-mir-196a-2	29331043

The top 1–25 miRNAs are shown in the second column while the top 26–50 are in the fifth column. As a result, 6 and 24 of the top 10 and top 50 were confirmed by the known experimental literature, respectively.

Definition 1 (Reachable Probability Matrix)

A reachable probability matrix for path $P = (A_1, A_2, \dots, A_{l+1})$ in a heterogeneous network can be defined as follows:

$$R_p = T_{A_1 A_2} T_{A_2 A_3} \cdots T_{A_l A_{l+1}}. \quad (\text{Equation 4})$$

Given two entities, $a \in SM$ and $b \in miRNA$, the relatedness between a and b can be calculated based on the HeteSim measure. Because HeteSim is a path-based relevance measurement, it is significant to consider different search paths that connect SM and miRNA. In general, the contribution of short paths may be more significant than long paths. Therefore, we only considered the paths from an SM to a miRNA in our heterogeneous network with length less than four (see Figure 4 step 2). If the length l of the search path is even, the search path can be divided into two parts with equal length. As we can see from Figure 4 step 2, a relevance path from an SM to a miRNA along a sequence of object types with length two can be indicated as $SM \xrightarrow{R_1} SM \xrightarrow{R_2} miRNA$. The search path $P = (SM, SM, miRNA)$ can be expressed as $P = (P_L P_R)$, where the left path $P_L = SM \xrightarrow{R_1} SM$ and the

right path $P_R = SM \xrightarrow{R_2} miRNA$. The reachable probability matrix for the left and right paths in the heterogeneous network are $R_{P_L} = T_{ss}$ and $R_{P_R} = T_{ms}$, respectively. Finally, the HeteSim score between a and b based on the path P can be calculated as follows:

$$HeteSim(a, b | P) = \frac{R_{P_L}(a, :) \times (R_{P_R}^{-1}(b, :))^T}{R_{P_L}(a, :)_2 \times R_{P_R}^{-1}(b, :)_2}, \quad (\text{Equation 5})$$

where P_R^{-1} is the reverse path of P_R . We give an example of using HeteSim to calculate the score between s_3 and m_1 under the path $P = (SM, SM, miRNA)$, i.e., $R_{P_L} = T_{ss}$ and $R_{P_R}^{-1} = T_{ms}$:

$$HeteSim(s_3, m_1 | P) = \frac{T_{ss}(3, :) \times (T_{ms}(1, :))^T}{T_{ss}(3, :)_2 \times T_{ms}(1, :)_2}. \quad (\text{Equation 6})$$

For the other three search paths with the length two listed in Figure 4 step 2, the corresponding HeteSim scores can also be calculated. If the length l of the search path is odd, the path cannot be equally divided

Table 4. Top 50 miRNAs Associated with 5-Aza-2'-Deoxycytidine Were Predicted by HSSMMA Based on Dataset 1

SM	miRNA	Evidence	SM	miRNA	Evidence
CID 451668	hsa-mir-125b-1	26198104	CID 451668	hsa-mir-16-1	26198104
CID 451668	hsa-mir-125b-2	26198104	CID 451668	hsa-mir-26a-1	unconfirmed
CID 451668	hsa-mir-19b-1	unconfirmed	CID 451668	hsa-mir-181b-1	unconfirmed
CID 451668	hsa-mir-24-1	26198104	CID 451668	hsa-mir-155	26198104
CID 451668	hsa-mir-18a	unconfirmed	CID 451668	hsa-mir-132	unconfirmed
CID 451668	hsa-mir-17	26198104	CID 451668	hsa-mir-181b-2	unconfirmed
CID 451668	hsa-mir-324	unconfirmed	CID 451668	hsa-mir-194-1	unconfirmed
CID 451668	hsa-mir-203a	26577858	CID 451668	hsa-let-7c	unconfirmed
CID 451668	hsa-mir-145	26198104	CID 451668	hsa-mir-346	unconfirmed
CID 451668	hsa-mir-21	26198104	CID 451668	hsa-mir-320a	26198104
CID 451668	hsa-mir-20a	26198104	CID 451668	hsa-mir-197	unconfirmed
CID 451668	hsa-mir-19a	26198104	CID 451668	hsa-mir-199a-2	unconfirmed
CID 451668	hsa-mir-23a	unconfirmed	CID 451668	hsa-mir-200c	23626803
CID 451668	hsa-mir-27a	26198104	CID 451668	hsa-mir-128-2	unconfirmed
CID 451668	hsa-let-7a-1	unconfirmed	CID 451668	hsa-let-7a-2	unconfirmed
CID 451668	hsa-mir-181a-1	26198104	CID 451668	hsa-let-7a-3	26227220
CID 451668	hsa-mir-124-1	unconfirmed	CID 451668	hsa-mir-126	26198104
CID 451668	hsa-mir-27b	26198104	CID 451668	hsa-mir-128-1	27705931
CID 451668	hsa-let-7b	26708866	CID 451668	hsa-mir-221	unconfirmed
CID 451668	hsa-mir-328	unconfirmed	CID 451668	hsa-mir-133a-1	unconfirmed
CID 451668	hsa-mir-124-2	23200812	CID 451668	hsa-mir-342	unconfirmed
CID 451668	hsa-mir-124-3	23200812	CID 451668	hsa-mir-205	unconfirmed
CID 451668	hsa-let-7d	26802971	CID 451668	hsa-mir-149	27783537
CID 451668	hsa-mir-181a-2	26198104	CID 451668	hsa-mir-1-2	unconfirmed
CID 451668	hsa-mir-24-2	26198104	CID 451668	hsa-mir-137	26198104

The top 1–25 miRNAs are shown in the second column while the top 26–50 are in the fifth column. As a result, 6 and 24 of the top 10 and top 50 were confirmed by the known experimental literature, respectively.

into two parts. For example, a relevance path from an SM to a miRNA along a sequence of object types with length three can be indicated as $SM \xrightarrow{R_1} SM \xrightarrow{R_2} SM \xrightarrow{R_3} miRNA$. Therefore, we consider the following two cases: (1) $P_L = SM \xrightarrow{R_1} SM \xrightarrow{R_2} SM$ and $P_R = SM \xrightarrow{R_3} miRNA$. The reachable probability matrix for the left and right paths can be defined as $R_{P_L} = T_{ss}T_{ss}$ and $R_{P_R} = T_{sm}$, respectively; (2) $P_L = SM \xrightarrow{R_1} SM$ and $P_R = SM \xrightarrow{R_2} SM \xrightarrow{R_3} miRNA$. The reachable probability matrix of the left and right paths are $R_{P_L} = T_{ss}$ and $R_{P_R} = T_{ss}T_{sm}$, respectively. The final HeteSim value of the search path $P = (SM, SM, SM, miRNA)$ would be obtained through calculating the average of two HeteSim values based on the above two different cases. We can also calculate HeteSim score of another search path with the length three listed in Figure 4 step 2.

Consequently, we would obtain six different HeteSim scores of six relevance paths and integrate these scores to obtain the final association scores between a and b , which can be defined as follows (see Figure 4 step 3):

$$S(a, b) = \sum_{l=2}^3 \left(\beta^{l-1} \times \sum_{P_i \in \psi_l} HeteSim(a, b | P_i) \right), \quad (\text{Equation 7})$$

where a is the entity of object type SM, b is the entity of object type miRNA, and ψ_l is the set of path P_i from SM to miRNA with the length of l . β is a decay factor, which can further punish the longer paths.

SUPPLEMENTAL INFORMATION

Supplemental Information includes four tables and can be found with this article online at <https://doi.org/10.1016/j.omtn.2018.12.002>.

AUTHOR CONTRIBUTIONS

All authors contributed important elements to the work presented herein. J.Q. implemented the experiments, analyzed the results, and wrote the paper. X.C. conceived the project, developed the prediction method, designed the experiments, analyzed the results, revised the paper, and supervised the project. Y.-Z.S., Y.Z., S.-B.C., Z.M.,

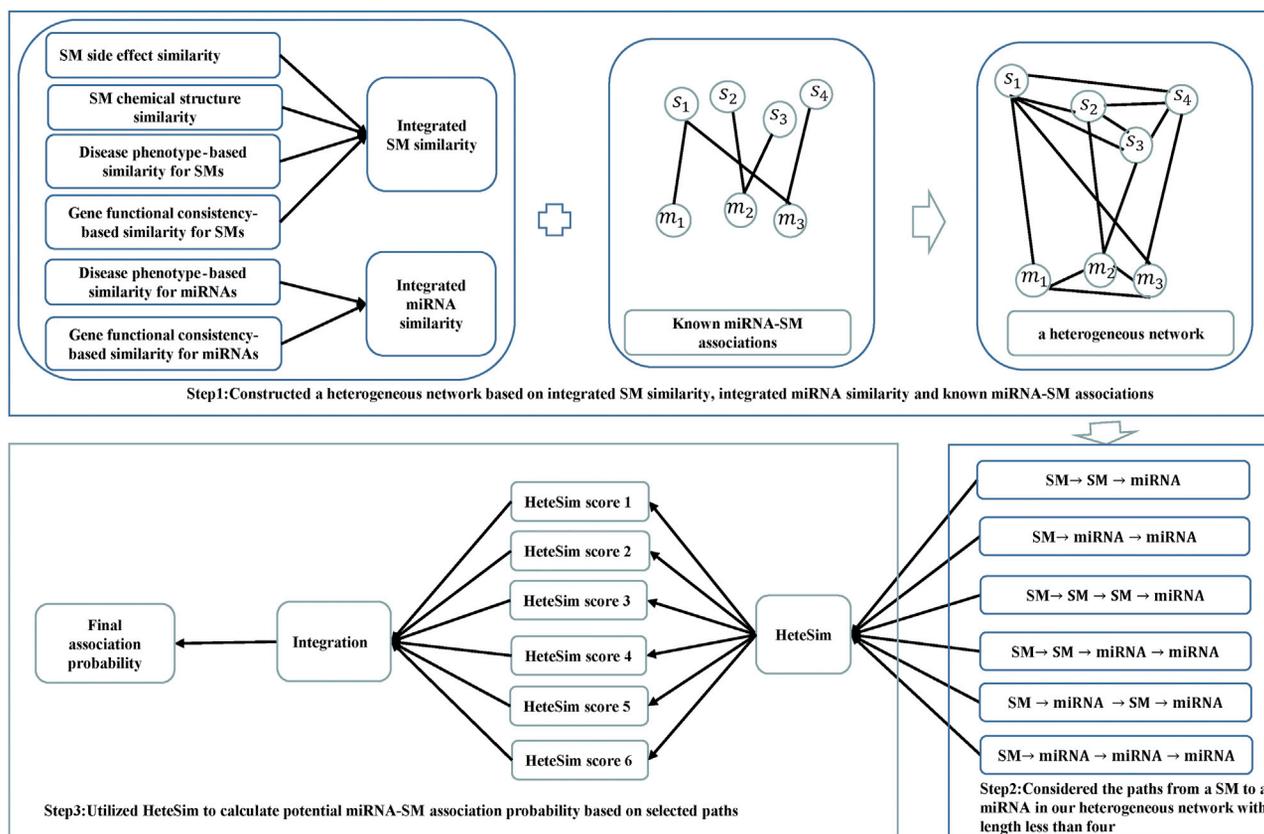


Figure 4. Flowchart of Potential miRNA-SM Association Prediction Based on HSSMMA

We constructed a miRNA-SM heterogeneous network based on known miRNA-SM associations, integrated SM similarity, and integrated miRNA similarity, and we obtained potential association probability of miRNA-SM associations by implementing HeteSim on the heterogeneous.

Z.-H.Y., and J.-Q.L. analyzed the results and revised the paper. All authors read and approved the final manuscript.

CONFLICTS OF INTEREST

The authors declare no competing interests.

ACKNOWLEDGMENTS

J.Q. was supported by Outstanding Innovation Scholarship for Doctoral Candidate of “Double First Rate” Construction Disciplines of CUMT.

REFERENCES

- Fu, X.D. (2014). Non-coding RNA: a new frontier in regulatory biology. *Natl. Sci. Rev.* 1, 190–204.
- Glinsky, G.V. (2008). Phenotype-defining functions of multiple non-coding RNA pathways. *Cell Cycle* 7, 1630–1639.
- Cooper, T.A., Wan, L., and Dreyfuss, G. (2009). RNA and disease. *Cell* 136, 777–793.
- Diederichs, S. (2012). Non-coding RNA and disease. *RNA Biol.* 9, 701–702.
- Chen, X., Sun, Y.-Z., Zhang, D.-H., Li, J.-Q., Yan, G.-Y., An, J.-Y., and You, Z.H. (2017). NRDTD: a database for clinically or experimentally supported non-coding RNAs and drug targets associations. *Database (Oxford)* 2017, bax057.
- Mruk, K., and Chen, J.K. (2015). Thinking big with small molecules. *J. Cell Biol.* 209, 7–9.
- Cundliffe, E. (1987). On the nature of antibiotic binding sites in ribosomes. *Biochimie* 69, 863–869.
- Deigan, K.E., and Ferré-D’Amaré, A.R. (2011). Riboswitches: discovery of drugs that target bacterial gene-regulatory RNAs. *Acc. Chem. Res.* 44, 1329–1338.
- Bevilacqua, P.C., Ritchey, L.E., Su, Z., and Assmann, S.M. (2016). Genome-Wide Analysis of RNA Secondary Structure. *Annu. Rev. Genet.* 50, 235–266.
- Bartel, D.P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116, 281–297.
- Iwakawa, H.O., and Tomari, Y. (2015). The Functions of MicroRNAs: mRNA Decay and Translational Repression. *Trends Cell Biol.* 25, 651–665.
- Wen, D., Danquah, M., Chaudhary, A.K., and Mahato, R.I. (2015). Small molecules targeting microRNA for cancer therapy: Promises and obstacles. *J. Control. Release* 219, 237–247.
- Chen, X., Yan, C.C., Zhang, X., Zhang, X., Dai, F., Yin, J., and Zhang, Y. (2016). Drug-target interaction prediction: databases, web servers and computational models. *Brief. Bioinform.* 17, 696–712.
- Disney, M.D., Winkelsas, A.M., Velagapudi, S.P., Southern, M., Fallahi, M., and Childs-Disney, J.L. (2016). Inforna 2.0: A Platform for the Sequence-Based Design of Small Molecules Targeting Structured RNAs. *ACS Chem. Biol.* 11, 1720–1728.
- Costales, M.G., Haga, C.L., Velagapudi, S.P., Childs-Disney, J.L., Phinney, D.G., and Disney, M.D. (2017). Small Molecule Inhibition of microRNA-210 Reprograms an Oncogenic Hypoxic Circuit. *J. Am. Chem. Soc.* 139, 3446–3455.

16. Haga, C.L., Velagapudi, S.P., Strivelli, J.R., Yang, W.Y., Disney, M.D., and Phinney, D.G. (2015). Small Molecule Inhibition of miR-544 Biogenesis Disrupts Adaptive Responses to Hypoxia by Modulating ATM-mTOR Signaling. *ACS Chem. Biol.* *10*, 2267–2276.
17. Seth, P.P., Miyaji, A., Jefferson, E.A., Sannes-Lowery, K.A., Osgood, S.A., Propp, S.S., Ranken, R., Massire, C., Sampath, R., Ecker, D.J., et al. (2005). SAR by MS: discovery of a new class of RNA-binding small molecules for the hepatitis C virus: internal ribosome entry site IIA subdomain. *J. Med. Chem.* *48*, 7099–7102.
18. Parsons, J., Castaldi, M.P., Dutta, S., Dibrov, S.M., Wyles, D.L., and Hermann, T. (2009). Conformational inhibition of the hepatitis C virus internal ribosome entry site RNA. *Nat. Chem. Biol.* *5*, 823–825.
19. Carnevali, M., Parsons, J., Wyles, D.L., and Hermann, T. (2010). A modular approach to synthetic RNA binders of the hepatitis C virus internal ribosome entry site. *ChemBioChem* *11*, 1364–1367.
20. Chen, C.Z., Sobczak, K., Hoskins, J., Southall, N., Marugan, J.J., Zheng, W., Thornton, C.A., and Austin, C.P. (2012). Two high-throughput screening assays for aberrant RNA-protein interactions in myotonic dystrophy type 1. *Anal. Bioanal. Chem.* *402*, 1889–1898.
21. Connelly, C.M., Boer, R.E., Moon, M.H., Gareiss, P., and Schneekloth, J.S., Jr. (2017). Discovery of Inhibitors of MicroRNA-21 Processing Using Small Molecule Microarrays. *ACS Chem. Biol.* *12*, 435–443.
22. Shortridge, M.D., and Varani, G. (2015). Structure based approaches for targeting non-coding RNAs with small molecules. *Curr. Opin. Struct. Biol.* *30*, 79–88.
23. Lorenz, D.A., and Garner, A.L. (2016). A click chemistry-based microRNA maturation assay optimized for high-throughput screening. *Chem. Commun. (Camb.)* *52*, 8267–8270.
24. Chen, X., Ren, B., Chen, M., Wang, Q., Zhang, L., and Yan, G. (2016). NLLSS: Predicting Synergistic Drug Combinations Based on Semi-supervised Learning. *PLoS Comput. Biol.* *12*, e1004975.
25. Jamal, S., Perival, V., and Scaria, V.; Open Source Drug Discovery Consortium (2012). Computational analysis and predictive modeling of small molecule modulators of microRNA. *J. Cheminform.* *4*, 16.
26. Zhang, S., Chen, L., Jung, E.J., and Calin, G.A. (2010). Targeting microRNAs with small molecules: from dream to reality. *Clin. Pharmacol. Ther.* *87*, 754–758.
27. Chen, X., Guan, N.N., Sun, Y.Z., Li, J.Q., and Qu, J. (2018). MicroRNA-small molecule association identification: from experimental results to computational models. *Brief. Bioinform.* Published online October 16, 2018. <https://doi.org/10.1093/bib/bby098>.
28. Wang, J., Meng, F., Dai, E., Yang, F., Wang, S., Chen, X., Yang, L., Wang, Y., and Jiang, W. (2016). Identification of associations between small molecule drugs and miRNAs based on functional similarity. *Oncotarget* *7*, 38658–38669.
29. Meng, F., Dai, E., Yu, X., Zhang, Y., Chen, X., Liu, X., Wang, S., Wang, L., and Jiang, W. (2013). Constructing and characterizing a bioactive small molecule and microRNA association network for Alzheimer's disease. *J. R. Soc. Interface* *11*, 20131057.
30. Jiang, W., Chen, X., Liao, M., Li, W., Lian, B., Wang, L., Meng, F., Liu, X., Chen, X., Jin, Y., and Li, X. (2012). Identification of links between small molecules and miRNAs in human cancers based on transcriptional responses. *Sci. Rep.* *2*, 282.
31. Lv, Y., Wang, S., Meng, F., Yang, L., Wang, Z., Wang, J., Chen, X., Jiang, W., Li, Y., and Li, X. (2015). Identifying novel associations between small molecules and miRNAs based on integrated molecular networks. *Bioinformatics* *31*, 3638–3644.
32. Li, J., Lei, K., Wu, Z., Li, W., Liu, G., Liu, J., Cheng, F., and Tang, Y. (2016). Network-based identification of microRNAs as potential pharmacogenomic biomarkers for anticancer drugs. *Oncotarget* *7*, 45584–45596.
33. Qu, J., Chen, X., Sun, Y.Z., Li, J.Q., and Ming, Z. (2018). Inferring potential small molecule-miRNA association based on triple layer heterogeneous network. *J. Cheminform.* *10*, 30.
34. Chen, X., Xie, D., Zhao, Q., and You, Z.H. (2017). MicroRNAs and complex diseases: from experimental results to computational models. *Brief. Bioinform.* Published online October 17, 2017. <https://doi.org/10.1093/bib/bbx130>.
35. Chen, X., Wang, L., Qu, J., Guan, N.N., and Li, J.Q. (2018). Predicting miRNA-disease association based on inductive matrix completion. *Bioinformatics* *34*, 4256–4265.
36. Chen, X., Xie, D., Wang, L., Zhao, Q., You, Z.H., and Liu, H. (2018). BNPMDA: Bipartite Network Projection for miRNA-Disease Association prediction. *Bioinformatics* *34*, 3178–3186.
37. Chen, X., Yin, J., Qu, J., and Huang, L. (2018). MDHGI: Matrix Decomposition and Heterogeneous Graph Inference for miRNA-disease association prediction. *PLoS Comput. Biol.* *14*, e1006418.
38. Chen, X., and Huang, L. (2017). LRSSLMDA: Laplacian Regularized Sparse Subspace Learning for miRNA-Disease Association prediction. *PLoS Comput. Biol.* *13*, e1005912.
39. Shi, C., Kong, X., Huang, Y., Yu, P.S., and Wu, B. (2014). HeteSim: A General Framework for Relevance Measure in Heterogeneous Networks. *IEEE Trans. Knowl. Data Eng.* *26*, 2479–2492.
40. Liu, X., Wang, S., Meng, F., Wang, J., Zhang, Y., Dai, E., Yu, X., Li, X., and Jiang, W. (2013). SM2miR: a database of the experimentally validated small molecules' effects on microRNA expression. *Bioinformatics* *29*, 409–411.
41. Liu, Y., Dong, Z., Liang, J., Guo, Y., Guo, X., Shen, S., Kuang, G., and Guo, W. (2016). Methylation-mediated repression of potential tumor suppressor miR-203a and miR-203b contributes to esophageal squamous cell carcinoma development. *Tumour Biol.* *37*, 5621–5632.
42. Zhang, Y., Cheng, C., He, D., Shi, W., Fu, C., Wang, X., and Zeng, C. (2016). Transcriptional gene silencing of dopamine D3 receptor caused by let-7d mimics in immortalized renal proximal tubule cells of rats. *Gene* *580*, 89–95.
43. Shan, Z.N., Tian, R., Zhang, M., Gui, Z.H., Wu, J., Ding, M., Zhou, X.F., and He, J. (2016). miR128-1 inhibits the growth of glioblastoma multiforme and glioma stem-like cells via targeting BMI1 and E2F3. *Oncotarget* *7*, 78813–78826.
44. Zhu, X.W., Yao, D.M., Wu, D.H., Wen, X.M., Yang, J., Guo, H., Yang, L., Deng, Z.Q., Zhang, Y.Y., Qian, W., et al. (2016). Let-7a-3 hypomethylation is associated with favorable/intermediate karyotypes but not with survival in acute myeloid leukemia. *Tumour Biol.* *37*, 491–501.
45. Windle, R., Bell, P.R., and Shaw, D. (1987). Five year results of a randomized trial of adjuvant 5-fluorouracil and levamisole in colorectal cancer. *Br. J. Surg.* *74*, 569–572.
46. Airley, R. (2009). *Cancer Chemotherapy: Basic Science to the Clinic* (Wiley).
47. Andreuccetti, M., Allegrini, G., Antonuzzo, A., Malvaldi, G., Conte, P.F., Danesi, R., Del Tacca, M., and Falcone, A. (1996). Azidothymidine in combination with 5-fluorouracil in human colorectal cell lines: in vitro synergistic cytotoxicity and DNA-induced strand-breaks. *Eur. J. Cancer* *32A*, 1219–1226.
48. Longley, D.B., Harkin, D.P., and Johnston, P.G. (2003). 5-fluorouracil: mechanisms of action and clinical strategies. *Nat. Rev. Cancer* *3*, 330–338.
49. Rossi, L., Bonmassar, E., and Faraoni, I. (2007). Modification of miR gene expression pattern in human colon cancer cells following exposure to 5-fluorouracil in vitro. *Pharmacol. Res.* *56*, 248–253.
50. Hernández-Vargas, H., Ballestar, E., Carmona-Saez, P., von Kobbe, C., Bañón-Rodríguez, I., Esteller, M., Moreno-Bueno, G., and Palacios, J. (2006). Transcriptional profiling of MCF7 breast cancer cells in response to 5-Fluorouracil: relationship with cell cycle changes and apoptosis, and identification of novel targets of p53. *Int. J. Cancer* *119*, 1164–1175.
51. Shah, M.Y., Pan, X., Fix, L.N., Farwell, M.A., and Zhang, B. (2011). 5-Fluorouracil drug alters the microRNA expression profiles in MCF-7 breast cancer cells. *J. Cell. Physiol.* *226*, 1868–1878.
52. Huang, S., Wu, B., Li, D., Zhou, W., Deng, G., Zhang, K., and Li, Y. (2014). Knockdown of astrocyte elevated gene-1 inhibits tumor growth and modifies microRNAs expression profiles in human colorectal cancer cells. *Biochem. Biophys. Res. Commun.* *444*, 338–345.
53. Wang, T., Huang, B., Guo, R., Ma, J., Peng, C., Zu, X., Tang, H., and Lei, X. (2015). A let-7b binding site SNP in the 3'-UTR of the Bcl-xL gene enhances resistance to 5-fluorouracil and doxorubicin in breast cancer cells. *Oncol. Lett.* *9*, 1907–1911.
54. Han, J., Liu, Z., Wang, N., and Pan, W. (2016). MicroRNA-874 inhibits growth, induces apoptosis and reverses chemoresistance in colorectal cancer by targeting X-linked inhibitor of apoptosis protein. *Oncol. Rep.* *36*, 542–550.
55. Yao, Y., Chen, S., Zhou, X., Xie, L., and Chen, A. (2014). 5-FU and ixabepilone modify the microRNA expression profiles in MDA-MB-453 triple-negative breast cancer cells. *Oncol. Lett.* *7*, 541–547.

56. Simpson, E., and Santen, R.J. (2015). Celebrating 75 years of oestradiol. *J. Mol. Endocrinol.* *55*, T1–T20.
57. Sathish Kumar, B., Kumar, A., Singh, J., Hasanain, M., Singh, A., Fatima, K., Yadav, D.K., Shukla, V., Luqman, S., Khan, F., et al. (2014). Synthesis of 2-alkoxy and 2-benzyloxy analogues of estradiol as anti-breast cancer agents through microtubule stabilization. *Eur. J. Med. Chem.* *86*, 740–751.
58. Aqil, F., Jeyabalan, J., Munagala, R., Singh, I.P., and Gupta, R.C. (2016). Prevention of hormonal breast cancer by dietary jamun. *Mol. Nutr. Food Res.* *60*, 1470–1481.
59. Li, S., Hu, R., Wang, C., Guo, F., Li, X., and Wang, S. (2014). miR-22 inhibits proliferation and invasion in estrogen receptor α -positive endometrial endometrioid carcinomas cells. *Mol. Med. Rep.* *9*, 2393–2399.
60. Zhang, C.L., Wang, H., Yan, C.Y., Gao, X.F., and Ling, X.J. (2017). Deregulation of RUNX2 by miR-320a deficiency impairs steroidogenesis in cumulus granulosa cells from polycystic ovary syndrome (PCOS) patients. *Biochem. Biophys. Res. Commun.* *482*, 1469–1476.
61. Rao, X., Di Leva, G., Li, M., Fang, F., Devlin, C., Hartman-Frey, C., Burow, M.E., Ivan, M., Croce, C.M., and Nephew, K.P. (2011). MicroRNA-221/222 confers breast cancer fulvestrant resistance by regulating multiple signaling pathways. *Oncogene* *30*, 1082–1097.
62. Patra, S.K., and Bettuzzi, S. (2009). Epigenetic DNA-(cytosine-5-carbon) modifications: 5-aza-2'-deoxycytidine and DNA-demethylation. *Biochemistry (Mosc.)* *74*, 613–619.
63. McCabe, M.T., Low, J.A., Daignault, S., Imperiale, M.J., Wojno, K.J., and Day, M.L. (2006). Inhibition of DNA methyltransferase activity prevents tumorigenesis in a mouse model of prostate cancer. *Cancer Res.* *66*, 385–392.
64. Issa, J.P., Garcia-Manero, G., Giles, F.J., Mannari, R., Thomas, D., Faderl, S., Bayar, E., Lyons, J., Rosenfeld, C.S., Cortes, J., and Kantarjian, H.M. (2004). Phase 1 study of low-dose prolonged exposure schedules of the hypomethylating agent 5-aza-2'-deoxycytidine (decitabine) in hematopoietic malignancies. *Blood* *103*, 1635–1640.
65. Hurtubise, A., and Momparler, R.L. (2004). Evaluation of antineoplastic action of 5-aza-2'-deoxycytidine (Dacogen) and docetaxel (Taxotere) on human breast, lung and prostate carcinoma cell lines. *Anticancer Drugs* *15*, 161–167.
66. Xu, W.Q., Huang, Y.M., and Xiao, H.F. (2015). [Expression Analysis and Epigenetics of MicroRNA let-7b in Acute Lymphoblastic Leukemia]. *Zhongguo Shi Yan Xue Ye Xue Za Zhi* *23*, 1535–1541.
67. Shimizu, T., Suzuki, H., Nojima, M., Kitamura, H., Yamamoto, E., Maruyama, R., Ashida, M., Hatahira, T., Kai, M., Masumori, N., et al. (2013). Methylation of a panel of microRNA genes is a novel biomarker for detection of bladder cancer. *Eur. Urol.* *63*, 1091–1100.
68. Manavalan, T.T., Teng, Y., Litchfield, L.M., Muluhngwi, P., Al-Rayyan, N., and Klinge, C.M. (2013). Reduced expression of miR-200 family members contributes to antiestrogen resistance in LY2 human breast cancer cells. *PLoS ONE* *8*, e62334.
69. Ghasemi, A., Fallah, S., and Ansari, M. (2016). MicroRNA-149 is epigenetically silenced tumor-suppressive microRNA, involved in cell proliferation and downregulation of AKT1 and cyclin D1 in human glioblastoma multiforme. *Biochem. Cell Biol.* *94*, 569–576.
70. Garzon, R., Marcucci, G., and Croce, C.M. (2010). Targeting microRNAs in cancer: rationale, strategies and challenges. *Nat. Rev. Drug Discov.* *9*, 775–789.
71. Gottlieb, A., Stein, G.Y., Ruppin, E., and Sharan, R. (2011). PREDICT: a method for inferring novel drug indications with application to personalized medicine. *Mol. Syst. Biol.* *7*, 496.
72. Lv, S., Li, Y., Wang, Q., Ning, S., Huang, T., Wang, P., Sun, J., Zheng, Y., Liu, W., Ai, J., and Li, X. (2012). A novel method to quantify gene set functional association based on gene ontology. *J. R. Soc. Interface* *9*, 1063–1072.
73. Hattori, M., Okuno, Y., Goto, S., and Kanehisa, M. (2003). Development of a chemical structure comparison method for integrated analysis of chemical and genomic information in the metabolic pathways. *J. Am. Chem. Soc.* *125*, 11853–11865.
74. Kuhn, M., Campillos, M., Letunic, I., Jensen, L.J., and Bork, P. (2010). A side effect resource to capture phenotypic effects of drugs. *Mol. Syst. Biol.* *6*, 343.
75. Knox, C., Law, V., Jewison, T., Liu, P., Ly, S., Frolkis, A., Pon, A., Banco, K., Mak, C., Neveu, V., et al. (2011). DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res.* *39*, D1035–D1041.
76. Zhu, F., Shi, Z., Qin, C., Tao, L., Liu, X., Xu, F., Zhang, L., Song, Y., Liu, X., Zhang, J., et al. (2012). Therapeutic target database update 2012: a resource for facilitating target-oriented drug discovery. *Nucleic Acids Res.* *40*, D1128–D1136.
77. Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* *40*, D109–D114.
78. Friedman, R.C., Farh, K.K., Burge, C.B., and Bartel, D.P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* *19*, 92–105.
79. Li, Y., Qiu, C., Tu, J., Geng, B., Yang, J., Jiang, T., and Cui, Q. (2014). HMDD v2.0: a database for experimentally supported human microRNA and disease associations. *Nucleic Acids Res.* *42*, D1070–D1074.
80. Jiang, Q., Wang, Y., Hao, Y., Juan, L., Teng, M., Zhang, X., Li, M., Wang, G., and Liu, Y. (2009). miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res.* *37*, D98–D104.
81. Ruepp, A., Kowarsch, A., Schmidl, D., Buggenthin, F., Brauner, B., Dunger, I., Fobo, G., Frishman, G., Montrone, C., and Theis, F.J. (2010). PhenomiR: a knowledgebase for microRNA expression in diseases and biological processes. *Genome Biol.* *11*, R6.
82. Zheng, X., Wang, Y., Tian, K., Zhou, J., Guan, J., Luo, L., and Zhou, S. (2017). Fusing multiple protein-protein similarity networks to effectively predict lncRNA-protein interactions. *BMC Bioinformatics* *18* (Suppl 12), 420.
83. Zeng, X., Liao, Y., Liu, Y., and Zou, Q. (2017). Prediction and Validation of Disease Genes Using HeteSim Scores. *IEEE/ACM Trans. Comput. Biol. Bioinformatics* *14*, 687–695.
84. Xiao, Y., Zhang, J., and Deng, L. (2017). Prediction of lncRNA-protein interactions using HeteSim scores based on heterogeneous networks. *Sci. Rep.* *7*, 3664.