JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

Cite This: J. Chem. Inf. Model. 2019, 59, 1668−1679

pubs.acs.org/jcim

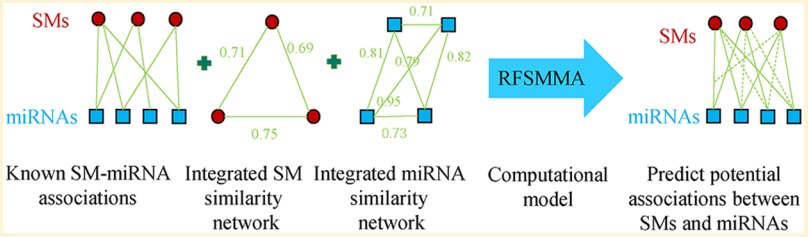
RFSMMA: A New Computational Model to Identify and Prioritize Potential Small Molecule−MiRNA Associations

Chun-Chun Wang,[†] Xing Chen,*[,†] Jia Qu,[†] Ya-Zhou Sun,[‡] and Jian-Qiang Li[‡]

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

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

Supporting Information



ABSTRACT: More and more studies found that many complex human diseases occur accompanied by aberrant expression of microRNAs (miRNAs). Small molecule (SM) drugs have been utilized to treat complex human diseases by affecting the expression of miRNAs. Several computational methods were proposed to infer underlying associations between SMs and miRNAs. In our study, we proposed a new calculation model of random forest based small molecule−miRNA association prediction (RFSMMA) which was based on the known SM−miRNA associations in the SM2miR database. RFSMMA utilized the similarity of SMs and miRNAs as features to represent SM−miRNA pairs and further implemented the machine learning algorithm of random forest to train training samples and obtain a prediction model. In RFSMMA, integrating multiple kinds of similarity can avoid the bias of single similarity and choosing more reliable features from original features can represent SM−miRNA pairs more accurately. We carried out cross validations to assess predictive accuracy of RFSMMA. As a result, RFSMMA acquired AUCs of 0.9854, 0.9839, 0.7052, and 0.9917 ± 0.0008 under global leave-one-out cross validation (LOOCV), miRNA-fixed local LOOCV, SM-fixed local LOOCV, and 5-fold cross validation, respectively, under data set 1. Based on data set 2, RFSMMA obtained AUCs of 0.8456, 0.8463, 0.6653, and 0.8389 ± 0.0033 under four cross validations according to the order mentioned above. In addition, we implemented a case study on three common SMs, namely, 5-fluorouracil, 17β-estradiol, and 5-aza-2′-deoxycytidine. Among the top 50 associated miRNAs of these three SMs predicted by RFSMMA, 31, 32, and 28 miRNAs were verified, respectively. Therefore, RFSMMA is shown to be an effective and reliable tool for identifying underlying SM−miRNA associations.

INTRODUCTION

RNA molecules play an important role in the transmission of genetic information. The Human Genome Project revealed that only about 1.5% of the genes in the human genome can encode proteins.[1] In human tissues, the amount of noncoding RNAs (ncRNAs) is three times higher than the amount of protein-coding RNAs.[2] NcRNAs are a large family including various categories of long (longer than 200 nucleotides) and short (less than 200 nucleotides) ncRNA molecules.[3,4] MicroRNAs (miRNAs) are short endogenously available ncRNAs with about 20 nucleotides. Its discovery dates back to 1993, when a small ncRNA transcript lin4 was reported in *C. elegans*.[5] At that time, it was unclear whether miRNAs have any biological function. The following research revealed that miRNAs are key regulatory elements of gene expression.[6] MiRNAs are essential mediators in variety of cellular processes and closely related to disease.[7,8]

MiRNAs regulate the expression level of their target genes through binding to the 3′ untranslated region of their target mRNAs.[9] Research demonstrated that human miRNAs regulate the expression of more than 60% of transcripts.[10] That means one miRNA may have more than one target mRNAs, which results in direct regulation of different proteins.[11,12] MiRNAs play many important roles in regulation of diverse cellular pathways.[13,14] The expression patterns of miRNAs are tissue specific and the abnormal expression often influences the status of cells.[15] Many clinical and experimental evidence demonstrated that miRNAs participate in various human complex diseases including cancer, cardiovascular, infectious, immune-related, and neurological disorders.[16,17] For example, Bommer et al.[18] found that miRNA34's expression level is significantly decreased in 40 percent of nonsmall cell lung cancers. Due to the key role of miRNAs in human complex diseases, there are numerous opportunities to identify miRNAs as potential therapeutic targets in drug development.

Received: February 12, 2019
Published: March 6, 2019

Downloaded via CHINESE UNIV OF HONG KONG SHENZHEN on May 18, 2024 at 05:19:53 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

© 2019 American Chemical Society

DOI: 10.1021/acs.jcim.9b00129
J. Chem. Inf. Model. 2019, 59, 1668−1679

In current, most targets of small molecule (SM) in clinical are proteins.[19] However, in all the expressed human proteins, only 10−15% are thought to be disease-related.[20] Moreover, in these disease-related proteins, many molecules lack distinctive motifs that SMs can bind with, meaning that they are not suitable to be targeted.[20] The implication is that designed small-molecule drugs based on human expressed proteins target a very small percentage of the human genome. Actually, there were only about 700 human gene products affected by presently approved drugs. This means that merely about 0.05% of the genome can be affected by drugs.[21] If ncRNAs could be targets of small molecules, the targets of small molecules would be highly abundant.

As one class of ncRNAs, miRNAs are well-validated SMs therapeutic targets. Some SMs can function as miRNA inhibitors. For example, genistein, an iosflavone, has been reported to inhibit miR-21 in renal cancer A498 cells and tumor xenografts.[22] Besides, losartan is an antihypertensive drug, which can suppress the expression of miR-503 and miR-181d in mice and further attenuates deterioration of renal tissue.[24] In addition, previous research discovered that matrine could reduce the expression of miR-21 in breast cancer MCF-7 cells.[23] Targapremir-210 can suppress the growth of miR-210 in breast cancer by binding to the Dicer site of immature miR-210.[25] Moreover, caffeic acid could be used to treat diabetic kidney disease of rats by inhibiting the expression of miR-342, miR-133b, and miR30a in rats.[26] Previous experiments validated that miR-345 and miR-7 are related to the resistance of cisplatin in breast cancer cells, on the contrary, miR-302b can make breast cancer cells sensitize to cisplatin.[27,28] Deep understanding of the relationship between miRNA and SMs will benefit not only the novel drug discovery but also the old drug positioning and drug resistance studies.

Recently, several approaches have been developed to design SM drugs targeting miRNAs. These approaches fall into three main categories.[29] Some studies carried out high-throughput to screen known compounds and found out SMs that can interact with the enzymes in miRNA biogenesis.[30] The platform cat-ELISA is one of these approaches for screening miRNA-SM interactions.[31] Another type of approach focuses on SMs that are predicted to interact with miRNA based on sequence or structure analysis.[32] The platform of inforna can identify miRNA-SM associations based on sequence.[32] Finally, other approaches combined bioinformatics analysis with a fluorescence-based screening system together.[33] These studies held great promise for identifying novel agents targeting miRNA and old drugs repositioning greatly. However, identification of the associations between SMs and miRNAs through biological experiments would cost a lot of time because of the diversity of biological systems and drug structures.[34] Thus, it is necessary to develop effective and accurate calculation models to infer the relationships between SMs and miRNAs to promote pharmacogenomics studies of human diseases.

Several computational models have been developed to identify the associations between SMs and miRNAs.[35] For example, Jiang et al.[36] developed a computational method to identify SM−miRNA associations in different human cancers. For an investigated cancer, they first identified cancer-related miRNAs (CRMs) and searched significant gene ontology modules (GOMs) for each CRM by judging whether differentially expressed targets of CRM were enriched in certain GOM according to the P-value obtained from hypergeometric test. In each GOM of a CRM, the target genes of the CRM were divided into overexpressed genes and underexpressed genes. Then they computed the Kolmogorov−Smirnov (KS) for genes of overexpression (KS_{up}) and genes of underexpression (KS_{down}) by considering the target genes' different expression under the CRM regulation and a SM treatment, which were integrated as similarity score to evaluate the similarity of genes with different expression between the CRM regulation and SM perturbation under this GOM. Finally, if the significant association (P-value ≤ 0.05) between the CRM and a SM was discovered in no less than 80 percent of significant GOMs of the CRM after implementing KS-test, the SM was considered to be associated with the CRM. Furthermore, Meng et al.[37] constructed and characterized a bioactive SM and miRNA relationship network for Alzheimer's disease. For the purpose of predicting potential associations between SM and Alzheimer's disease related miRNA (ADM), they implemented KS-test by considering the signatures of SMs and ADMs. For each ADM, they calculated similarity score of differentially expressed genes between the ADM regulation and SM perturbation and regarded the significant SM−ADM pairs (P-value ≤ 0.01) as potential associations. It worth noting that SM−ADM associations could be divided to two categories of positive and negative associations according to the signs of similarity scores. Distinguishing different associations might help to develop anti-miRNA drugs and miRNA mimic drugs. Besides, Lv et al.[38] established a heterogeneous molecular network through integrating similarity networks of SM and miRNA and SM−miRNA association network. Then they extended the traditional algorithm of random walk with restart (RWR) through carrying out iterative formula on the heterogeneous network with two types of nodes rather than partial network only involving single type node to predict potential miRNA targets for SMs. Thus, this method could be utilized to infer the underlying associated-miRNAs for SM without considering whether the SM contains known associated miRNAs. Recently, Li et al.[39] proposed a SM−miRNA network-based inference (SMiR-NBI) model. In SMiR-NBI, they connected SMs and their associated miRNAs to construct a heterogeneous network based on known databases. Then a network-based inference (NBI) algorithm was utilized to identity new SM−miRNA associations. Briefly, the initial resources of investigated SM situated in related miRNAs of the SM. Then the resources of each associated miRNA were equally distributed to their adjacent SMs. Next, the SMs in the last step distributed resources to their adjacent miRNAs. The final resource scores of miRNAs represent their possibility associated with the SM. In more recent, Wang et al.[40] presented an approach to infer potential SM−miRNA associations by considering the functional similarity of genes with different expression following drug action and miRNA disturbance. Besides, Qu et al.[41] developed an approach of a triple layer heterogeneous network based SM−miRNA association prediction (TLHNSMMA) for identifying underlying SM−miRNA associations. In their model, based on SM and miRNA's information, disease information was introduced as bridge to construct a triple layer heterogeneous network. Then based on the triple layer network, they constructed two iterative formulas to infer potential SM−miRNA relationships and miRNA−disease relationships until the two iterative formulas converge. Furthermore, Guan et al.[42] proposed an inference model to identity SM−miRNA association based on graphlet interaction. They established similarity networks of SM and

DOI: 10.1021/acs.jcim.9b00129
J. Chem. Inf. Model. 2019, 59, 1668−1679

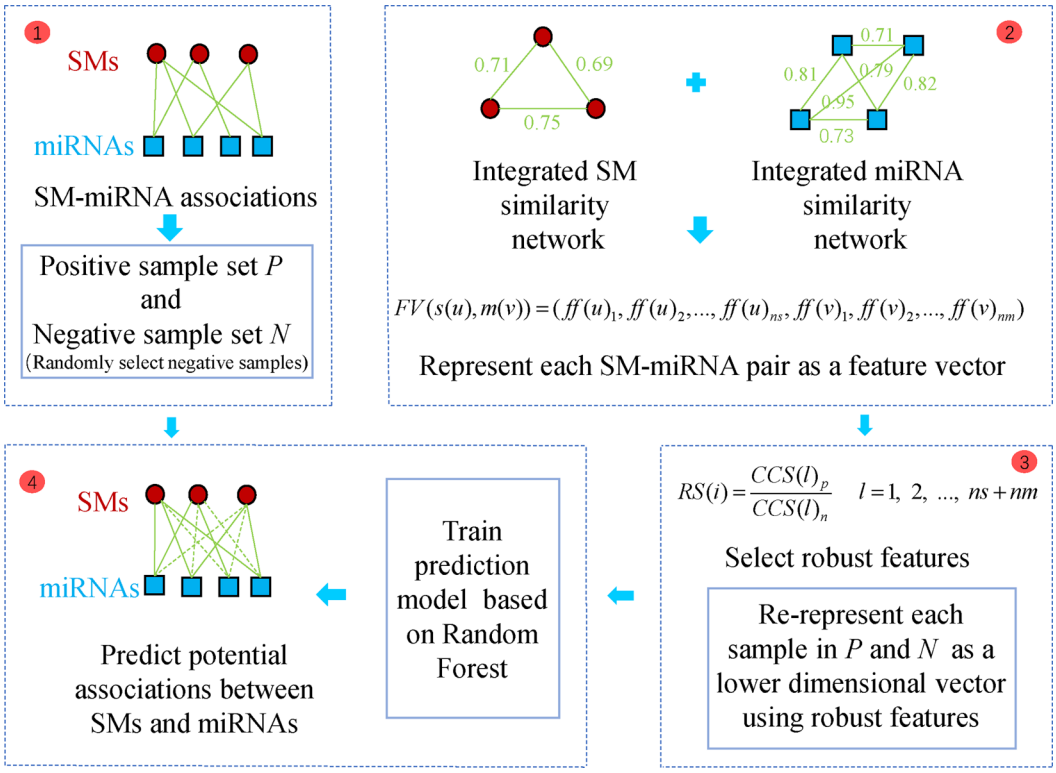


Figure 1. Flowchart of RFSMMA model to infer underlying associations between SMs and miRNAs.

miRNA, then they defined 28 kinds of graphlet interaction isomers. From the perspective of miRNA, they count the numbers of each isomer of two miRNA nodes based on miRNA similarity network, which were used to calculated SM−miRNA association score. Similarly, they also calculated SM−miRNA association score from SM perspective. The final SM−miRNA association score was the average of the above two scores.

In our study, we proposed an effective calculated method of random forest based small molecule−MiRNA association (RFSMMA) prediction. Actually, the algorithm of random forest has been employed to predict the binding of in vivo transcription factor,[43] identify enhancers[44] and peptides,[45] infer miRNA−disease associations,[46] and work for other biological issues.[47,48] Here we combined SM side effect similarity, SM chemical structure similarity, disease phenotype-based similarity of SM, gene functional consistency-based similarity of SM and disease phenotype-based similarity of miRNA, and gene functional consistency-based similarity of miRNA. The integrated similarities of SM and miRNA were used as features of SM and miRNA in the model of RFSMMA, respectively. And the machine learning technology of random forest was utilized to infer underlying relationships between SMs and miRNAs. Besides, to efficiently identify potential associations, a filter-based method was employed for extracting reliable features of SMs and miRNAs. It is worth noting that we established two kinds of data sets in which known SM−miRNA associations are same, but the numbers of SMs and miRNAs are different. The known associations between SMs and miRNAs were downloaded from the database of SM2miR v1.0.[49] Data set 1 contained all SMs and miRNAs collected from several databases, while data set 2 only contained a part of SMs and miRNAs with known associations. To assess the prediction accuracy of RFSMMA, we conducted global and local leave-one-out cross validation (LOOCV) along with 5-fold cross validation in both data sets 1 and 2 and inspected the area under the receiver operating characteristics curve (AUC). In data set 1, RFSMMA achieved AUC of 0.9854 under global LOOCV. Besides, the AUC under SM-fixed and miRNA-fixed local LOOCV were 0.7502 and 0.9839 respectively. In addition, RFSMMA obtained average AUC and standard deviation of 0.9917 ± 0.0008 under 5-fold cross validation. In data set 2, RFSMMA obtained AUCs of 0.8456, 0.6653, 0.8463, and 0.8389 ± 0.0033 under different cross validations implemented in data set 1. We also carried out case studies in three different SMs of 5-fluorouracil (5-FU), 17β-estradiol (E2), and 5-aza-2′-deoxycytidine (5-AZA-CdR) based on data set 1. Specifically, RFSMMA was utilized to infer underlying miRNAs associated with investigate SM. It is worth noting that the three different SMs were considered as novel SMs when we implemented RFSMMA. At last, among the top 50 potential miRNAs predicted by RFSMMA for 5-FU, E2, and 5-AZA-CdR, there were 31, 32, 28 miRNAs confirmed by published references, respectively. Thus, we believe that RFSMMA would be an effective tool to infer SM−miRNA associations for further experimental verification.

MATERIALS AND METHODS

Small Molecule−miRNA Associations. We collected 664 known SM−miRNA associations from SM2miR v1.0.[49] These known associations; 831 SMs collected from SM2miR, DrugBank,[50] and PubChem;[51] and 541 miRNAs obtained from databases of SM2miR, miR2Disease,[52] HMDD,[53] and PhenomiR[54] make up the data set 1. However, in data set 1, there are some SMs and miRNAs without any association information. Therefore, we removed these SMs and miRNAs. Then, we obtained data set 2 which consists of 664 known SM−miRNA associations, 39 SMs, and 286 miRNAs. To

DOI: 10.1021/acs.jcim.9b00129
J. Chem. Inf. Model. 2019, 59, 1668−1679

better describe relationships between SMs and miRNAs, we defined the adjacency matrix A. The element $A(i, j)$ is equal to 1 if SM $s(i)$ was verified to be associated with miRNA $m(j)$; otherwise, it is 0. Furthermore, we utilized variable ns and nm to denote the number of SMs and miRNAs, respectively.

Integrated SM Similarity. For avoiding the bias of single SM similarity, we integrated four types of similarity for SM, namely, disease phenotype-based similarity of SM,[55] gene functional consistency-based similarity of SM,[56] SM chemical structure-based similarity,[57] and SM side effect-based similarity[55] according to the similarity calculation and similarity integration method in previous study.[38] First, we utilized matrix S_S^d to express the information on disease phenotype-based similarity of SM in which the element $S_S^d(i, j)$ denotes the similarity between SM $s(i)$ and SM $s(j)$ based on disease phenotype. Besides, we employed matrices S_S^g, S_S^c, and S_S^s to express SM similarity based on gene functional consistency, chemical structure, and side effect, respectively. Then, the four types of similarity were combined as integrated similarity as follows:

$$S_S = \frac{\lambda_1 S_S^d + \lambda_2 S_S^g + \lambda_3 S_S^c + \lambda_4 S_S^s}{\lambda_1 + \lambda_2 + \lambda_3 + \lambda_3} \quad (1)$$

where the default value λ_1, λ_2, λ_3, and λ_4 were set as 1. S_S is the integrated SM similarity matrix. The element $S_S(i, j)$ of S_S represents the integrated similarity between SM $s(i)$ and SM $s(j)$.

Integrated miRNA Similarity. We combined two kinds of similarity for miRNA, namely, disease phenotype-based similarity of miRNA[55] and gene functional consistency-based similarity of miRNA.[56] First, we used matrices S_M^d and S_M^g to express the information on miRNA similarity based on disease phenotype and gene functional consistency, respectively. Then the two kinds of similarity for miRNA can be integrated in a same approach utilized in previous study[38] as follows:

$$S_M = \frac{\mu_1 S_M^d + \mu_2 S_M^g}{\mu_1 + \mu_2} \quad (2)$$

where the default value μ_1 and μ_2 were set as 1. And S_M is the integrated miRNA similarity matrix The element $S_M(i, j)$ of S_M represents the integrated similarity between miRNAs $m(i)$ and $m(j)$.

RFSMMA. In this study, we established the prediction model of RFSMMA to infer underlying SM−miRNA associations, which was motivated from previous study.[58] The process of establishing prediction model was roughly segmented into four steps (See Figure 1): 1. Select positive and negative samples to form positive and negative sample set for training. 2. Represent each SM−miRNA pair as a feature vector. 3. Choose a robust feature subset and rerepresent each SM−miRNA pair as a low dimensional vector. 4. Train positive and negative samples represented by feature vectors based on random forest algorithm and obtain a final prediction model. The details of every step will be described in the following.

First, we established positive and negative sample sets for training. The positive sample set P contained all 664 known associations recorded in the database of SM2miR v1.0. Based on the ratio of 1:1, the negative sample set N also included 664 negative samples. In this study, the negative samples were randomly selected from unlabeled samples.

Second, each SM−miRNA pair was denoted by a feature vector. We constructed the prediction model based on the hypothesis that similar SMs are more likely to be related with similar miRNAs and vice versa. Thus, based on biological information on SM and miRNA, we calculated similarities of SM and miRNA which were combined as feature vector to denote training sample. Training samples were divided into positive class and negative class which were assigned different labels. Based on the training samples and machine learning algorithm of random forest, we predict SM−miRNA relationships by inspecting the similarities between investigated SM and other SMs and the similarities between investigated miRNA and other miRNAs. Therefore, the similarity features we utilized could contribute to the prediction of SM−miRNA associations.

The integrated similarities of SM and miRNA were utilized as features in this study. Specifically, SM contained ns features $(f_1, f_2, ..., fx, ..., f_{ns})$ in which the xth feature f_x denotes the integrated similarity between investigated SM and the xth SM. And an SM $s(u)$ was denoted by a ns-dimensional vector as follows:

$$FV(s(u)) = (f(u)_1, f(u)_2, ..., f(u)_x..., f(u)_{ns}) \quad (3)$$

where the xth element of $FV(s(u))$ is the integrated similarity between SM $s(x)$ and SM $s(u)$.

For a miRNA $m(v)$, an nm-dimensional vector was utilized to represent it. Similarly, the nm-dimensional vector was defined as follows:

$$FV(m(v)) = (f(v)_1, f(v)_2, ..., f(v)_y..., f(v)_{nm}) \quad (4)$$

where the yth element of $FV(m(v))$ is the integrated similarity between miRNAs $m(y)$ and $m(v)$.

Then we stitched vectors $FV(s(u))$ and $FV(m(v))$ together to denote the SM−miRNA pair composed of $s(u)$ and $m(v)$ as a $(ns + nm)$-dimensional feature vector:

$$FV(s(u), m(v)) = (f(u)_1, f(u)_2, ..., f(u)_{ns}, f(v)_1, f(v)_2, ..., f(v)_{nm}) \quad (5)$$

Then we normalized each feature value $f(u)_x$ and $f(v)_y$ to $ff(u)_x$ and $ff(v)_y$ as below:

$$ff(u)_x = \frac{f(u)_x - f_{\min}}{f_{\max} - f_{\min}} \quad x = 1, 2, ..., ns \quad (6)$$

$$ff(v)_y = \frac{f(v)_y - f_{\min}}{f_{\max} - f_{\min}} \quad y = 1, 2, ..., nm \quad (7)$$

where $f_{\max}$ and $f_{\min}$ are the maximum and the minimum among all $f(u)_x$ $(x = 1, 2, ..., ns)$ and $f(v)_y$ $(y = 1, 2, ..., nm)$.

Third, to more efficiently distinguish associated SM−miRNA pairs from unassociated SM−miRNA pairs, we selected more robust features and redenoted SM−miRNA pairs by lower dimensional feature vectors. The more robust feature means that the feature value is higher in the sample set of P and lower in the sample set of N or higher in N and lower in P. Here, a filter-based technique was employed to achieve our purpose.

SM−miRNA pairs in the positive sample set P and negative sample set N have been represented by feature vectors after the second step. We utilized FV_k^P (or FV_k^N) to express the feature vector of the kth sample in P (or N). Besides, $CS(l)_p$ and

DOI: 10.1021/acs.jcim.9b00129
J. Chem. Inf. Model. 2019, 59, 1668−1679

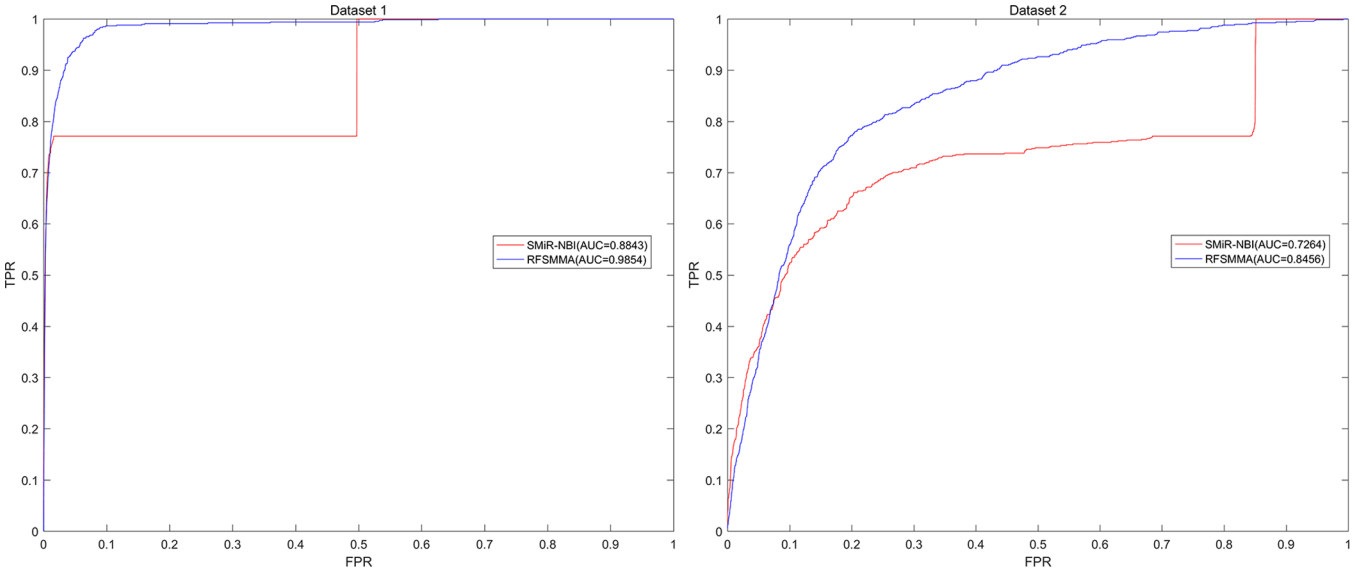


Figure 2. Comparison between the performances of RFSMMA and SMiR-NBI according to the ROC curve and the value of AUC under global LOOCV. As a result, RFSMMA achieved AUCs of 0.9854 and 0.8456 which are higher than the AUCs of 0.8843 and 0.7264 obtained from SMiR-NBI based on data sets 1 and 2.

$CS(l)_n$ were used to represent the cumulative score of the lth feature in P and N respectively. Therefore, $CS(l)_p$ and $CS(l)_n$ were defined as below:

$$CS(l)_p = \sum_{k=1}^{|P|} FV_k^P(l) \tag{8}$$

$$CS(l)_n = \sum_{k=1}^{|N|} FV_k^N(l) \tag{9}$$

where $|P|$ and $|N|$ are the sample numbers in P and N respectively. $FV_k^P(l)$ denotes the lth feature value of FV_k^P, meanwhile $FV_k^N(l)$ expresses the lth feature value of FV_k^N.

$CS(l)_p$ and $CS(l)_n$ were further normalized to $CCS(l)_p$ and $CCS(l)_n$ in a similar way to eqs 6 and 7. Then we employed $RS(l)$ to represent the robust score of the lth feature, which can be computed as below:

$$RS(l) = \frac{CCS(l)_p}{CCS(l)_n} \quad l = 1, 2, ..., ns + nm \tag{10}$$

where $CCS(l)_p$ and $CCS(l)_n$ reflect the enrichment of feature value of the lth feature in P and N. The more robust feature we expected should be that the feature value is higher in P and lower in N or higher in N and lower in P. Thus, the most robust feature should have the highest or lowest robust score. In data set 1, the original dimension of feature vector is $ns_1 + nm_1$ (ns_1 and nm_1 express the number of SMs and miRNAs in data set 1, respectively). We would select k_1 robust features to represent samples in data set 1. According to previous study,[59] they selected 300 features from more than 3000 features to represent drug−target pairs as feature vectors. Our original feature dimension was more than 1000 so we set k_1 as 100 in our study. Specifically, we selected 100 more robust features from $ns_1 + nm_1$ features, where 60 features were selected from ns_1 features of SM and 40 features were selected from nm_1 features of miRNA roughly based on the original proportion of SM's feature and miRNA's feature number. Among the 60 features of SM, half of them possess the top 30 robust scores and half possess the last 30 robust scores of all ns_1 robust scores $RS(l)$ (l = 1, 2, ..., ns_1). For the 40 features of miRNA, half of them possess the top 20 robust scores and half possess the last 20 robust scores of all nm_1 robust scores $RS(l)$ ($l = ns_1 + 1, ns_1 + 2, ..., ns_1 + nm_1$). In data set 2, we selected k_2 features. Since the original feature dimension is relatively low, we simply set k_2 as 80 which is about one-fourth of the original feature dimension to ensure the sufficiency and effectiveness of the features. Similarly, we chose 80 more robust features from $ns_2 + nm_2$ (ns_2 denotes the number of SMs in data set 2, and nm_2 denotes the number of miRNAs in Data set 2) features. And 10 features were chosen from ns_2 features of SM and 70 features were chosen from nm_2 features of miRNA. After selecting robust feature subset, each SM−miRNA was denoted by a lower dimensional feature vector by utilizing these robust features.

Finally, we gave each positive sample a label of 1 and each negative sample a label of 0. Both positive and negative samples represented by feature vectors and labels were trained through Random-Forest-Regressor, an algorithm package based on random forest.[60] The candidate SM−miRNA pairs were scored by RFSMMA and the higher score of SM−miRNA pair denotes the higher probability that the SM and miRNA are associated with each other. It is worth noting that main parameters of Random-Forest-Regressor, n_estimators, max_features, and min_samples_leaf, were set to 100, 0.2, and 10 based on experience.

RESULTS

Performance Evaluation. We introduced global and local LOOCV as well as 5-fold cross validation to assess the prediction accuracy of RFSMMA under two different data sets. And the 664 known SM−miRNA associations included in these two data sets were downloaded from SM2miR v1.0 database.[49] Besides, the 664 known associations between SMs and miRNAs were also utilized in another previous method of SMiR-NBI[39] to infer potential SM−miRNA associations. Therefore, we compared the performance of RFSMMA with

DOI: 10.1021/acs.jcim.9b00129
J. Chem. Inf. Model. 2019, 59, 1668−1679

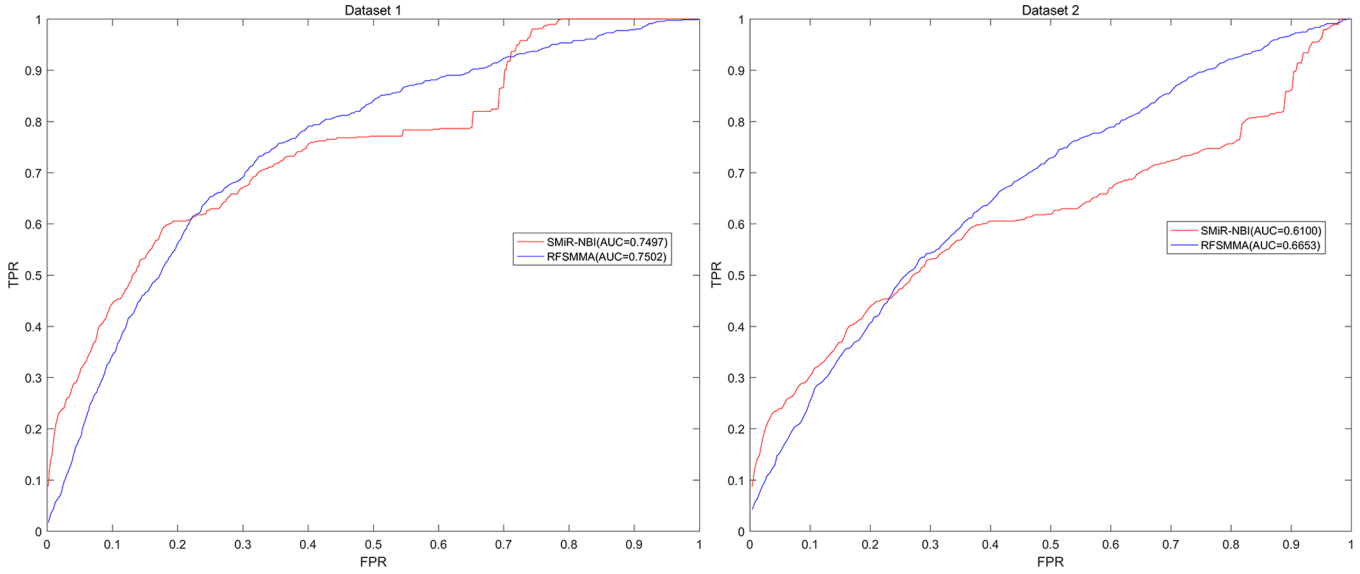


Figure 3. Comparison between performances of RFSMMA and SMiR-NBI according to the ROC curve and the value of AUC under SM-fixed local LOOCV. As a result, RFSMMA achieved AUCs of 0.7502 and 0.6653 which are higher than the AUCs of 0.7497 and 0.6100 obtained from SMiR-NBI based on data sets 1 and 2.

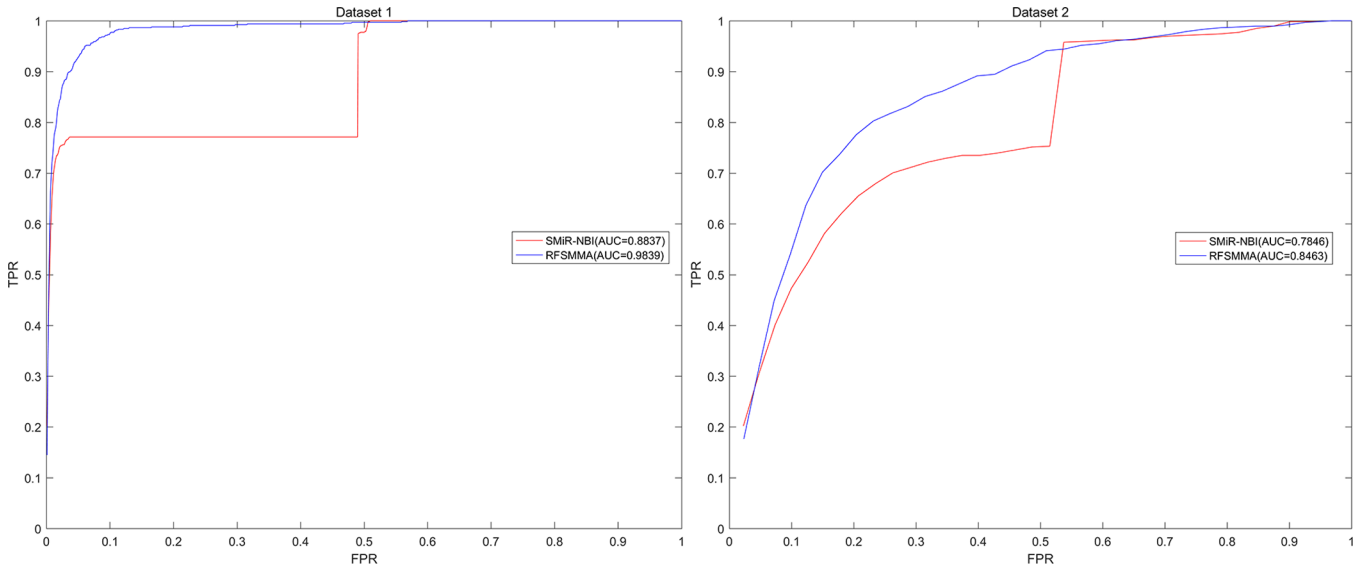


Figure 4. Comparison between performances of RFSMMA and SMiR-NBI according to the ROC curve and the value of AUC under miRNA-fixed local LOOCV. As a result, RFSMMA achieved AUCs of 0.9839 and 0.8463 which are higher than the AUCs of 0.8837 and 0.7846 obtained from SMiR-NBI based on data sets 1 and 2.

SMiR-NBI through their AUCs under different cross validations.

In global LOOCV, positive samples (associated SM−miRNA pair) take turns as a test sample while other positive samples were used as positive training samples. RFSMMA gives a score to the test sample and all unlabeled samples (unknown SM−miRNA pairs). Then we could obtain the ranking of the test sample after sorting all scores of the test sample and all unlabeled samples in decreasing order. Finally, the receiver operating characteristic (ROC) curves can be drawn in which the *X*-coordinate is false positive rate (FPR, 1 − specificity), and the *Y*-coordinate is the true positive rate (TPR, sensitivity). The points on ROC curves are TPRs against FPRs in different thresholds. Sensitivity indicates the proportion of test samples whose rankings were higher than the given threshold; meanwhile, specificity reveals the ratio of negative samples whose rankings were lower than the given threshold. We considered AUC as an evaluation criterion. And the higher AUC indicates more excellent prediction accuracy of prediction model. It can be seen from Figure 2 that RFSMMA achieved AUCs of 0.9854 and 0.8456 under global LOOCV based on data sets 1 and 2 which are higher than the AUCs of 0.8843 and 0.7264 obtained by SMiR-NBI. In local LOOCV, the distinction from global LOOCV is that the score of test sample was compared with scores of unlabeled samples just including the specific SM (or miRNA) contained in the test sample in each loop. Specifically, in SM-fixed local LOOCV, the score of test sample was ranked with all unlabeled samples which contain the SM involved in the test sample. While, in miRNA-fixed local LOOCV, the score of test

DOI: 10.1021/acs.jcim.9b00129
J. Chem. Inf. Model. 2019, 59, 1668−1679

Table 1. Top 50 Associated miRNAs of 5-FU Predicted by RFSMMA[a]

SM	miRNA	evidence	SM	miRNA	evidence
CID 3385	hsa-mir-218-1	unconfirmed	CID 3385	hsa-mir-335	unconfirmed
CID 3385	hsa-mir-181c	unconfirmed	CID 3385	hsa-mir-128-2	26198104
CID 3385	hsa-mir-181a-1	unconfirmed	CID 3385	hsa-mir-34c	unconfirmed
CID 3385	hsa-mir-155	28347920	CID 3385	hsa-mir-128-1	26198104
CID 3385	hsa-mir-152	26198104	CID 3385	hsa-mir-376a-1	unconfirmed
CID 3385	hsa-mir-151a	26198104	CID 3385	hsa-mir-339	unconfirmed
CID 3385	hsa-mir-28	unconfirmed	CID 3385	hsa-mir-186	unconfirmed
CID 3385	hsa-mir-146a	28466779	CID 3385	hsa-mir-24-1	26198104
CID 3385	hsa-mir-142	23619912	CID 3385	hsa-mir-218-2	28918035
CID 3385	hsa-mir-141	26198104	CID 3385	hsa-mir-214	unconfirmed
CID 3385	hsa-mir-140	19734943	CID 3385	hsa-mir-212	unconfirmed
CID 3385	hsa-mir-139	27173050	CID 3385	hsa-mir-222	26198104
CID 3385	hsa-mir-137	25550779	CID 3385	hsa-mir-210	26198104
CID 3385	hsa-mir-135b	26198104	CID 3385	hsa-mir-21	26198104
CID 3385	hsa-mir-183	27249600	CID 3385	hsa-mir-205	24396484
CID 3385	hsa-mir-376c	unconfirmed	CID 3385	hsa-mir-204	27095441
CID 3385	hsa-mir-135a-1	unconfirmed	CID 3385	hsa-mir-200c	26198104
CID 3385	hsa-mir-132	26198104	CID 3385	hsa-mir-1-1	unconfirmed
CID 3385	hsa-mir-130b	24119443	CID 3385	hsa-mir-1-2	unconfirmed
CID 3385	hsa-mir-130a	unconfirmed	CID 3385	hsa-mir-200b	26198104
CID 3385	hsa-mir-302b	26457704	CID 3385	hsa-mir-200a	24510588
CID 3385	hsa-mir-30d	28207045	CID 3385	hsa-mir-9-1	unconfirmed
CID 3385	hsa-mir-30e	26198104	CID 3385	hsa-mir-199b	unconfirmed
CID 3385	hsa-mir-31	26198104	CID 3385	hsa-mir-199a-2	26198104
CID 3385	hsa-mir-129-1	unconfirmed	CID 3385	hsa-mir-106a	26198104

[a]Here, 5-FU was considered as a novel SM which has no known associated-miRNAs when training the prediction model. The top 1−25 miRNAs are listed in the second column, and the top 26−50, in the fifth. As can be seen, 6 out of the top 10 and 31 out of the top 50 were validated according to the literature, respectively.

sample was ranked with all unlabeled samples which contain the miRNA involved in the test sample. Similarly, we can draw the ROC curves under these two kinds of local LOOCV. As shown in Figure 3 (under SM-fixed local LOOCV), RFSMMA achieved AUCs of 0.7502 and 0.6653 respectively based on data sets 1 and 2 which are higher than the AUCs of 0.7497 and 0.6100 of SMiR-NBI. In addition, as shown in Figure 4 (under miRNA-fixed local LOOCV), RFSMMA achieved AUCs of 0.9839 and 0.8463 while SMiR-NBI achieved AUCs of 0.8837 and 0.7846 based on different data sets.

In the 5-fold cross validation, positive samples were separated into five sets randomly and equally. Each of them would be left out as test set in turns meanwhile the remaining four sets were utilized for training prediction model. The sample in the test set and all unlabeled samples would be scored by RFSMMA. Then we found out the ranking of each test sample in scores including the test sample's score and all unlabeled samples' scores which were ranked in descending order. We repeated this procedure for 100 times to reduce the influence of performance difference due to samples divisions. Finally, RFSMMA achieved average AUC and standard deviation of 0.9917 and 0.0008 while SMiR-NBI obtained average AUC and standard deviation of 0.8554 and 0.0063 under data set 1. Based on data set 2, RFSMMA obtained AUC of 0.8389 ± 0.0033 and SMiR-NBI obtained AUC of 0.7104 ± 0.0087.

Case Studies. To further evaluate the prediction accuracy of RFSMMA utilized to infer underlying associated-miRNA for novel SM, we carried out case studies based on data set 1 on three common SMs, namely, 5-FU, E2, and 5-AZA-CdR. The novel SM means that we still do not discover known miRNAs associated with it. Thus, when 5-FU (E2 or 5-AZA-CdR) was considered as the investigated object of the case study, all known 5-FU−miRNA (E2−miRNA or 5-AZA-CdR−miRNA) associations in data set 1 were regarded as unlabeled pairs in the process of training prediction model of RFSMMA. Then RFSMMA was utilized to infer potential associated miRNAs for the SM. Finally, we observed how many miRNAs were confirmed by experiment in the top 50 list predicted by RFSMMA.

5-FU is a drug extensively used to treat cancer.[61] It exerts effects through inhibiting thymidylate synthase and incorporating into DNA or RNA.[62] We took 5-FU as the first investigated object of case study, and the predicted associated miRNAs for 5-FU were ranked according to their scores obtained from RFSMMA in descending order. At last, there were 6 confirmed associations between 5-FU and the top 10 inferred miRNAs according to the SM2miR database and previous literature works, in which 3 pairs are corroborated by the SM2miR database[49] and the other 3 pairs were confirmed by experimental literature. Among the pairs between 5-FU and the top 50 inferred miRNAs, there were 31 verified associations, in which 17 associations were verified by the SM2miR database[49] and 14 associations were validated from experimental literature (See Table 1). For example, real-time polymerase chain reaction confirmed that the expression level of miR-155 which ranks fourth in the top 50 inferred miRNAs was observably relevant to overall survival of patients after treatment with 5-FU in oropharyngeal carcinoma (OPSCC).[63] Besides, the overexpression of the eighth predicted 5-FU-related miRNA miR-146a in colorectal cancer HT-29 cells increased the cell proliferation through a method of dose-

DOI: 10.1021/acs.jcim.9b00129
J. Chem. Inf. Model. 2019, 59, 1668−1679

Table 2. Top 50 Associated miRNAs of E2 Inferred by RFSMMA[a]

SM	miRNA	evidence	SM	miRNA	evidence
CID 5757	hsa-mir-124-3	26198104	CID 5757	hsa-mir-128-1	26198104
CID 5757	hsa-mir-124-2	26198104	CID 5757	hsa-mir-1-2	26198104
CID 5757	hsa-mir-124-1	26198104	CID 5757	hsa-mir-128-2	26198104
CID 5757	hsa-mir-25	unconfirmed	CID 5757	hsa-mir-182	28678802
CID 5757	hsa-mir-92a-1	unconfirmed	CID 5757	hsa-mir-150	unconfirmed
CID 5757	hsa-mir-338	22996663	CID 5757	hsa-mir-195	unconfirmed
CID 5757	hsa-mir-19b-1	unconfirmed	CID 5757	hsa-mir-181b-2	unconfirmed
CID 5757	hsa-mir-27b	26198104	CID 5757	hsa-mir-1-1	26198104
CID 5757	hsa-mir-19b-2	unconfirmed	CID 5757	hsa-mir-139	unconfirmed
CID 5757	hsa-mir-27a	26198104	CID 5757	hsa-mir-146b	29331043
CID 5757	hsa-mir-148a	25928008	CID 5757	hsa-mir-486	unconfirmed
CID 5757	hsa-mir-181b-1	unconfirmed	CID 5757	hsa-mir-26b	24735615
CID 5757	hsa-mir-16-2	unconfirmed	CID 5757	hsa-mir-9-1	26198104
CID 5757	hsa-mir-18a	24245576	CID 5757	hsa-mir-135b	23143558
CID 5757	hsa-mir-17	26198104	CID 5757	hsa-mir-302b	26198104
CID 5757	hsa-mir-16-1	unconfirmed	CID 5757	hsa-mir-204	29789714
CID 5757	hsa-mir-125b-1	unconfirmed	CID 5757	hsa-mir-223	27884727
CID 5757	hsa-mir-30c-1	26198104	CID 5757	hsa-mir-9-2	26198104
CID 5757	hsa-mir-30a	29331043	CID 5757	hsa-mir-19a	29416771
CID 5757	hsa-mir-424	26198104	CID 5757	hsa-mir-378a	unconfirmed
CID 5757	hsa-mir-142	unconfirmed	CID 5757	hsa-mir-93	23492819
CID 5757	hsa-mir-125b-2	unconfirmed	CID 5757	hsa-mir-23b	26198104
CID 5757	hsa-mir-30c-2	26198104	CID 5757	hsa-mir-200c	26198104
CID 5757	hsa-mir-34a	24050776	CID 5757	hsa-mir-30b	26198104
CID 5757	hsa-mir-92a-2	unconfirmed	CID 5757	hsa-mir-10a	unconfirmed

[a]It is worth noting that, when training the prediction model, all known associations between E2 and miRNAs were removed. The top 1−25 miRNAs are listed in the second column, and the top 26−50, in the fifth. As can be seen, 6 out of the top 10 and 32 out of the top 50 were confirmed according to the literature, respectively.

dependent followed treatment with 5-FU, suggesting that miR-146a can enhance drug resistance to 5-FU.[64] Furthermore, the reduction of miR-142-3p levels which ranks ninth in the top inferred miRNAs can raise the sensitivity of SW1116 carcinoma of colon cells to 5-FU,[65] while the overexpression of miR-140 which ranks 11th causes chemoresistance to the action of 5-FU in HCT-116 carcinoma of colon cells.[66] In addition, ectopic expression of the twelfth predicted 5-FU-related miRNA miR-139-5p can make cells of LoVo and HCT-116 sensitize to 5-FU by targeting the NOTCH-1 gene and increasing 5-FU-induced apoptosis, suggesting a prospective remedial target for the antimultidrug resistance treatment of carcinoma of colon.[67]

E2 is a primary estrogen that plays important roles in the reproductive functions of females.[68] Recently, researchers have discovered that E2 is also involved in various biological responses, which can change male and female physiology.[68] Consequently, E2 is closely related to many human diseases and applied clinically.[69] After training prediction model of RFSMMA based on data set 1, we inferred associated miRNAs for E2. As a result, there were 6 verified associations between E2 and top 10 inferred miRNAs according to database and previous literatures, in which 5 associations were confirmed by the SM2miR database[49] and the other 1 association was confirmed by experimental literature. Among the pairs between E2 and the top 50 inferred miRNAs, there are 32 verified associations, in which 19 associations were authenticated by SM2miR database[49] and 13 associations were verified by experimental literature (See Table 2). Based on experimental evidence from references, after exposure to 100 nM E2 for 2 days, the level of miR-338-3p which is the sixth predicted E2-related miRNA was decreased by approximately half in islets and cell lines which secrete insulin.[70] The downregulation of miR-338-3p is related with expansion of the β cell of pregnant women, suggesting new therapeutic strategies to treat diabetes during pregnancy.[70] Besides, in the triple-negative breast carcinoma (TNBC) cell, E2 treatment significantly upregulated the lncRNA HOTAIR through suppressing miR-148a which ranks 11th among the top 50 predicted E2-related miRNAs, which may be an important regulation mechanism in cancer and provide novel targets for TNBC therapy.[71] Meanwhile, the 14th predicted E2-related miRNA miR-18a was significantly upregulated about 13-fold under the influence of E2 intervention in breast cancer tissues.[72] Furthermore, in the uterus and endometrial carcinoma, microarray analysis identified that miR-30a which ranks 19th was downregulated by E2.[73]

5-AZA-CdR is an underlying DNA methylation inhibitor.[74] It can incorporate into methylated CpG sites during DNA replication and inactivate DNA methyl transferase 1 (DNMT1) to take the antineoplastic action.[74] 5-AZA-CdR has been utilized to treat many carcinomas especially myeloid leukemia.[75,76] After carrying out case study for 5-AZA-CdR, we obtained the result as follows. Among the pairs between 5-AZA-CdR and the top 10 inferred miRNAs, there are 9 verified associations, in which 6 associations were confirmed by the SM2miR database[49] and 3 associations were confirmed by experimental literature. Besides, among the pairs between 5-AZA-CdR and the top 50 inferred miRNAs, there are 28 verified associations, in which 20 associations were verified by the SM2miR database[49] and 8 associations were confirmed according to experimental literature (See Table 3). Here we

DOI: 10.1021/acs.jcim.9b00129
J. Chem. Inf. Model. 2019, 59, 1668−1679

Table 3. Top 50 Associated miRNAs of Novel SM 5-AZA-CdR Predicted by RFSMMA[a]

SM	miRNA	evidence	SM	miRNA	evidence
CID 451668	hsa-mir-126	26198104	CID 451668	hsa-mir-30c-1	unconfirmed
CID 451668	hsa-mir-107	26198104	CID 451668	hsa-mir-142	unconfirmed
CID 451668	hsa-let-7e	22053057	CID 451668	hsa-mir-18a	unconfirmed
CID 451668	hsa-mir-19a	26198104	CID 451668	hsa-mir-22	26198104
CID 451668	hsa-let-7b	26708866	CID 451668	hsa-mir-92a-1	unconfirmed
CID 451668	hsa-let-7f-1	26198104	CID 451668	hsa-mir-199b	24659709
CID 451668	hsa-let-7i	26198104	CID 451668	hsa-let-7a-3	26227220
CID 451668	hsa-mir-20a	26198104	CID 451668	hsa-let-7c	unconfirmed
CID 451668	hsa-mir-195	23333942	CID 451668	hsa-mir-200c	23626803
CID 451668	hsa-mir-221	unconfirmed	CID 451668	hsa-mir-1-1	unconfirmed
CID 451668	hsa-mir-150	unconfirmed	CID 451668	hsa-mir-181a-1	26198104
CID 451668	hsa-mir-204	unconfirmed	CID 451668	hsa-mir-106a	unconfirmed
CID 451668	hsa-let-7a-1	unconfirmed	CID 451668	hsa-mir-27a	26198104
CID 451668	hsa-let-7d	26802971	CID 451668	hsa-mir-145	26198104
CID 451668	hsa-mir-17	26198104	CID 451668	hsa-mir-125b-2	26198104
CID 451668	hsa-mir-125a	26198104	CID 451668	hsa-mir-143	28391715
CID 451668	hsa-mir-10b	unconfirmed	CID 451668	hsa-mir-148a	unconfirmed
CID 451668	hsa-mir-106b	26198104	CID 451668	hsa-mir-16-2	26198104
CID 451668	hsa-mir-23b	26198104	CID 451668	hsa-mir-98	unconfirmed
CID 451668	hsa-mir-19b-1	unconfirmed	CID 451668	hsa-mir-30a	unconfirmed
CID 451668	hsa-mir-205	unconfirmed	CID 451668	hsa-mir-15b	unconfirmed
CID 451668	hsa-mir-93	26198104	CID 451668	hsa-mir-100	unconfirmed
CID 451668	hsa-let-7a-2	unconfirmed	CID 451668	hsa-mir-16-1	26198104
CID 451668	hsa-mir-125b-1	26198104	CID 451668	hsa-mir-139	unconfirmed
CID 451668	hsa-let-7g	26198104	CID 451668	hsa-mir-26a-1	unconfirmed

[a]Here, all known associations between 5-AZA-CdR and miRNAs were removed when training the prediction model. The top 1−25 miRNAs are shown in the second column, while the top 26−50, in the fifth. As can been seen, 9 out of the top 10 and 28 out of the top 50 were validated by the literature, respectively.

showed several associations between 5-AZA-CdR and miRNAs confirmed by experimental evidence from references. For example, after treatment with 10 nM 5-AZA-CdR, the third predicted associated-miRNA let-7e showed significant differences in embryos.[77] Besides, the qPCR experiment confirmed that 5-AZA-CdR could upregulate expression level of the fifth inferred associated-miRNA let-7b in cell lines MOLT-4 of acute lymphoblastic leukemia, indicating a new therapeutic target for ALL.[78] Furthermore, in gastric cancer cell line MGC-803, the expression of miR-195 which ranks ninth in the top 50 predicted miRNAs is upregulated after treated by 5-AZA-CdR.[79] Moreover, let-7d which ranks 4th can suppress dopamine D3 receptor (DRD3) via DNA methylation, while this suppression can be eliminated abolished by 5-AZA-CdR.[80]

As shown above, the results of case studies still proved outstanding prediction accuracy of RFSMMA. Thus, RFSMMA was utilized to predict potential miRNAs for all the 831 SMs in our study and all prediction result can be seen in Table S1 (see the Supporting Information).

DISCUSSION

MiRNAs play an indispensable role in the development of cells and numerous biological processes. Aberrantly expressed miRNAs have been demonstrated to be associated with many diseases and it would be significant if we could treat complex human diseases through targeting dysregulated miRNAs. Therefore, more and more studies focus on identifying relationships between SMs and miRNAs. Here, we proposed a calculation model of RFSMMA. We utilized known SM−miRNA associations to establish a positive sample set; meanwhile, we randomly obtained negative samples from unlabeled SM−miRNA pairs to form a negative sample set. To express the information on SM−miRNA pairs more comprehensively, we integrated four kinds of similarity of SM, two kinds of similarity of miRNA and denoted each SM−miRNA pair by a feature vector using integrated similarity of SM and miRNA. After training the positive and negative samples, RFSMMA can infer potential SM−miRNA associations. As mentioned above, RFSMMA would be a reliable computational method to identify potential relationships between SMs and miRNAs, which have been proved by the results of the cross validation and case study.

The satisfactory performance of RFSMMA benefits from several important factors. In the first place, the known SM−miRNA associations used in the model are reliable because they were obtained from very reliable database of SM2miR v1.0. Second, we employed various biological information on SMs and miRNAs to represent SM−miRNA pairs as feature vectors. Third, a filter-based method was employed for choosing effective features and cutting down the dimension of feature vectors. In such a manner, RFSMMA can precisely differentiate associated SM−miRNA pairs fromunassociated pairs. Fourthly, RFSMMA utilized the random forest algorithm which implements an unbiased estimator for generalization error. Therefore, our prediction model has a good generalization ability. Finally, RFSMMA could predict for novel SMs and miRNAs even though there were no known association information for these SMs and miRNAs.

However, there are also several weaknesses existing in our computational model. The training sample set of RFSMMA included both positive and negative sample sets. However, the negative SM−miRNA samples were hard or even impossible to

DOI: 10.1021/acs.jcim.9b00129
J. Chem. Inf. Model. 2019, 59, 1668−1679

be obtained. In this study, we randomly selected negative samples from unlabeled SM−miRNA samples to form negative sample set, which may influence the performance of our prediction model. Besides, known SM−miRNA associations are insufficient. Thus, we believe that our prediction model would obtain better accuracy as more SM−miRNA associations can be verified and more reliable negative SM−miRNA sample can be selected through data analysis in the future. In addition, we plan to combine more biological data to predict SM−miRNA associations. Combining multiple distance measures could reduce the bias of each similarity measurement. In our model, we utilized the similarity of SM and miRNA as features to construct feature vectors for each SM−miRNA pair. Thus, more reliable similarity of SM and miRNA would make feature vectors better represent SM−miRNA pairs. In the future, we will further integrate more biological information into this computational model such as miRNA−disease associations,[81−84] lncRNA−miRNA interactions,[85] and drug−target interactions.[86] In addition, predicting potential SM−lncRNA associations would also be an important research topic.[87−89] Finally, we would try to utilize the cancer hallmark network[90] to study associations between cancer drugs and miRNAs in the future.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.9b00129.

We utilized RFSMMA to rank all the candidate SM−miRNA pairs based on the known SM−miRNA associations in the SM2miR v1.0 database. The prediction result may contribute to further experimental verification or related research (XLSX)

■ AUTHOR INFORMATION

Corresponding Author

*Email: xingchen@amss.ac.cn.

ORCID

Xing Chen: 0000-0001-9028-5342

Funding

This work was supported by the National Natural Science Foundation of China (No. 61772531).

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Carninci, P.; Kasukawa, T.; Katayama, S.; Gough, J.; Frith, M. C.; Maeda, N.; Oyama, R.; Ravasi, T.; Lenhard, B.; Wells, C.; Kodzius, R.; Shimokawa, K.; Bajic, V. B.; Brenner, S. E.; Batalov, S.; Forrest, A. R.; Zavolan, M.; Davis, M. J.; Wilming, L. G.; Aidinis, V.; Allen, J. E.; Ambesi-Impiombato, A.; Apweiler, R.; Aturaliya, R. N.; Bailey, T. L.; Bansal, M.; Baxter, L.; Beisel, K. W.; Bersano, T.; Bono, H.; Chalk, A. M.; Chiu, K. P.; Choudhary, V.; Christoffels, A.; Clutterbuck, D. R.; Crowe, M. L.; Dalla, E.; Dalrymple, B. P.; de Bono, B.; Della Gatta, G.; di Bernardo, D.; Down, T.; Engstrom, P.; Fagiolini, M.; Faulkner, G.; Fletcher, C. F.; Fukushima, T.; Furuno, M.; Futaki, S.; Gariboldi, M.; Georgii-Hemming, P.; Gingeras, T. R.; Gojobori, T.; Green, R. E.; Gustincich, S.; Harbers, M.; Hayashi, Y.; Hensch, T. K.; Hirokawa, N.; Hill, D.; Huminiecki, L.; Iacono, M.; Ikeo, K.; Iwama, A.; Ishikawa, T.; Jakt, M.; Kanapin, A.; Katoh, M.; Kawasawa, Y.; Kelso, J.; Kitamura, H.; Kitano, H.; Kollias, G.; Krishnan, S. P.; Kruger, A.; Kummerfeld, S. K.; Kurochkin, I. V.; Lareau, L. F.; Lazarevic, D.; Lipovich, L.; Liu, J.; Liuni, S.; McWilliam, S.; Madan Babu, M.; Madera, M.; Marchionni, L.; Matsuda, H.; Matsuzawa, S.; Miki, H.; Mignone, F.; Miyake, S.; Morris, K.; Mottagui-Tabar, S.; Mulder, N.; Nakano, N.; Nakauchi, H.; Ng, P.; Nilsson, R.; Nishiguchi, S.; Nishikawa, S.; Nori, F.; Ohara, O.; Okazaki, Y.; Orlando, V.; Pang, K. C.; Pavan, W. J.; Pavesi, G.; Pesole, G.; Petrovsky, N.; Piazza, S.; Reed, J.; Reid, J. F.; Ring, B. Z.; Ringwald, M.; Rost, B.; Ruan, Y.; Salzberg, S. L.; Sandelin, A.; Schneider, C.; Schonbach, C.; Sekiguchi, K.; Semple, C. A.; Seno, S.; Sessa, L.; Sheng, Y.; Shibata, Y.; Shimada, H.; Shimada, K.; Silva, D.; Sinclair, B.; Sperling, S.; Stupka, E.; Sugiura, K.; Sultana, R.; Takenaka, Y.; Taki, K.; Tammoja, K.; Tan, S. L.; Tang, S.; Taylor, M. S.; Tegner, J.; Teichmann, S. A.; Ueda, H. R.; van Nimwegen, E.; Verardo, R.; Wei, C. L.; Yagi, K.; Yamanishi, H.; Zabarovsky, E.; Zhu, S.; Zimmer, A.; Hide, W.; Bult, C.; Grimmond, S. M.; Teasdale, R. D.; Liu, E. T.; Brusic, V.; Quackenbush, J.; Wahlestedt, C.; Mattick, J. S.; Hume, D. A.; Kai, C.; Sasaki, D.; Tomaru, Y.; Fukuda, S.; Kanamori-Katayama, M.; Suzuki, M.; Aoki, J.; Arakawa, T.; Iida, J.; Imamura, K.; Itoh, M.; Kato, T.; Kawaji, H.; Kawagashira, N.; Kawashima, T.; Kojima, M.; Kondo, S.; Konno, H.; Nakano, K.; Ninomiya, N.; Nishio, T.; Okada, M.; Plessy, C.; Shibata, K.; Shiraki, T.; Suzuki, S.; Tagami, M.; Waki, K.; Watahiki, A.; Okamura-Oho, Y.; Suzuki, H.; Kawai, J.; Hayashizaki, Y. The transcriptional landscape of the mammalian genome. *Science (New York, N.Y.)* **2005**, *309*, 1559−63.

(2) Hangauer, M. J.; Vaughn, I. W.; McManus, M. T. Pervasive transcription of the human genome produces thousands of previously unidentified long intergenic noncoding RNAs. *PLoS Genet.* **2013**, *9*, e1003569.

(3) Batista, P. J.; Chang, H. Y. Long noncoding RNAs: cellular address codes in development and disease. *Cell* **2013**, *152*, 1298−307.

(4) Kim, V. N.; Han, J.; Siomi, M. C. Biogenesis of small RNAs in animals. *Nat. Rev. Mol. Cell Biol.* **2009**, *10*, 126−39.

(5) Lee, R. C.; Feinbaum, R. L.; Ambros, V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **1993**, *75*, 843−54.

(6) Iwakawa, H. O.; Tomari, Y. The Functions of MicroRNAs: mRNA Decay and Translational Repression. *Trends Cell Biol.* **2015**, *25*, 651−65.

(7) Adams, B. D.; Parsons, C.; Walker, L.; Zhang, W. C.; Slack, F. J. Targeting noncoding RNAs in disease. *J. Clin. Invest.* **2017**, *127*, 761−771.

(8) Chen, X.; Xie, D.; Zhao, Q.; You, Z. H. MicroRNAs and complex diseases: from experimental results to computational models. *Briefings Bioinf.* **2017**, DOI: 10.1093/bib/bbx130.

(9) Fernandez-Valverde, S. L.; Taft, R. J.; Mattick, J. S. MicroRNAs in β-Cell Biology, Insulin Resistance, Diabetes and Its Complications. *Diabetes* **2011**, *60*, 1825−1831.

(10) Bartel, D. P. Metazoan MicroRNAs. *Cell* **2018**, *173*, 20−51.

(11) Faghihi, M. A.; Zhang, M.; Huang, J.; Modarresi, F.; Van der Brug, M. P.; Nalls, M. A.; Cookson, M. R.; St-Laurent, G., 3rd; Wahlestedt, C. Evidence for natural antisense transcript-mediated inhibition of microRNA function. *Genome Biol.* **2010**, *11*, R56.

(12) Lim, L. P.; Lau, N. C.; Garrett-Engele, P.; Grimson, A.; Schelter, J. M.; Castle, J.; Bartel, D. P.; Linsley, P. S.; Johnson, J. M. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **2005**, *433*, 769−73.

(13) O'Donnell, K. A.; Wentzel, E. A.; Zeller, K. I.; Dang, C. V.; Mendell, J. T. c-Myc-regulated microRNAs modulate E2F1 expression. *Nature* **2005**, *435*, 839−43.

(14) Viswanathan, S. R.; Powers, J. T.; Einhorn, W.; Hoshida, Y.; Ng, T. L.; Toffanin, S.; O'Sullivan, M.; Lu, J.; Phillips, L. A.; Lockhart, V. L.; Shah, S. P.; Tanwar, P. S.; Mermel, C. H.; Beroukhim, R.; Azam, M.; Teixeira, J.; Meyerson, M.; Hughes, T. P.; Llovet, J. M.; Radich, J.; Mullighan, C. G.; Golub, T. R.; Sorensen, P. H.; Daley, G. Q. Lin28 promotes transformation and is associated with advanced human malignancies. *Nat. Genet.* **2009**, *41*, 843−8.

(15) Suzuki, H.; Maruyama, R.; Yamamoto, E.; Kai, M. Epigenetic alteration and microRNA dysregulation in cancer. *Front. Genet.* **2013**, *4*, 258.

(16) Rupaimoole, R.; Slack, F. J. MicroRNA therapeutics: towards a new era for the management of cancer and other diseases. *Nat. Rev. Drug Discovery* **2017**, *16*, 203−222.
(17) Macfarlane, L. A.; Murphy, P. R. MicroRNA: Biogenesis, Function and Role in Cancer. *Curr. Genomics* **2010**, *11*, 537−61.
(18) Bommer, G. T.; Gerin, I.; Feng, Y.; Kaczorowski, A. J.; Kuick, R.; Love, R. E.; Zhai, Y.; Giordano, T. J.; Qin, Z. S.; Moore, B. B.; MacDougald, O. A.; Cho, K. R.; Fearon, E. R. p53-mediated activation of miRNA34 candidate tumor-suppressor genes. *Curr. Biol.* **2007**, *17*, 1298−307.
(19) Hopkins, A. L.; Groom, C. R. The druggable genome. *Nat. Rev. Drug Discovery* **2002**, *1*, 727−30.
(20) Dixon, S. J.; Stockwell, B. R. Identifying druggable disease-modifying gene products. *Curr. Opin. Chem. Biol.* **2009**, *13*, 549−55.
(21) Santos, R.; Ursu, O.; Gaulton, A.; Bento, A. P.; Donadi, R. S.; Bologa, C. G.; Karlsson, A.; Al-Lazikani, B.; Hersey, A.; Oprea, T. I.; Overington, J. P. A comprehensive map of molecular drug targets. *Nat. Rev. Drug Discovery* **2017**, *16*, 19−34.
(22) Zaman, M. S.; Shahryari, V.; Deng, G.; Thamminana, S.; Saini, S.; Majid, S.; Chang, I.; Hirata, H.; Ueno, K.; Yamamura, S.; Singh, K.; Tanaka, Y.; Tabatabai, Z. L.; Dahiya, R. Up-regulation of microRNA-21 correlates with lower kidney cancer survival. *PLoS One* **2012**, *7*, e31060.
(23) Li, L. Q.; Li, X. L.; Wang, L.; Du, W. J.; Guo, R.; Liang, H. H.; Liu, X.; Liang, D. S.; Lu, Y. J.; Shan, H. L.; Jiang, H. C. Matrine Inhibits Breast Cancer Growth Via miR-21/PTEN/Akt Pathway in MCF-7 Cells. *Cell. Physiol. Biochem.* **2012**, *30*, 631−641.
(24) Zhu, X.; Zhang, C.; Fan, Q.; Liu, X.; Yang, G.; Jiang, Y.; Wang, L. Inhibiting MicroRNA-503 and MicroRNA-181d with Losartan Ameliorates Diabetic Nephropathy in KKAy Mice. *Med. Sci. Monit.* **2016**, *22*, 3902−3909.
(25) Costales, M. G.; Haga, C. L.; Velagapudi, S. P.; Childs-Disney, J. L.; Phinney, D. G.; Disney, M. D. Small Molecule Inhibition of microRNA-210 Reprograms an Oncogenic Hypoxic Circuit. *J. Am. Chem. Soc.* **2017**, *139*, 3446−3455.
(26) Matboli, M.; Eissa, S.; Ibrahim, D.; Hegazy, M. G. A.; Imam, S. S.; Habib, E. K. Caffeic Acid Attenuates Diabetic Kidney Disease via Modulation of Autophagy in a High-Fat Diet/Streptozotocin-Induced Diabetic Rat. *Sci. Rep.* **2017**, *7*, 2263.
(27) Pogribny, I. P.; Filkowski, J. N.; Tryndyak, V. P.; Golubov, A.; Shpyleva, S. I.; Kovalchuk, O. Alterations of microRNAs and their targets are associated with acquired resistance of MCF-7 breast cancer cells to cisplatin. *Int. J. Cancer* **2010**, *127*, 1785−94.
(28) Cataldo, A.; Cheung, D. G.; Balsari, A.; Tagliabue, E.; Coppola, V.; Iorio, M. V.; Palmieri, D.; Croce, C. M. miR-302b enhances breast cancer cell sensitivity to cisplatin by regulating E2F1 and the cellular DNA damage response. *Oncotarget* **2016**, *7*, 786−97.
(29) Nguyen, D. D.; Chang, S. Development of Novel Therapeutic Agents by Inhibition of Oncogenic MicroRNAs. *Int. J. Mol. Sci.* **2017**, *19*, e65.
(30) Di Giorgio, A.; Tran, T. P.; Duca, M. Small-molecule approaches toward the targeting of oncogenic miRNAs: roadmap for the discovery of RNA modulators. *Future Med. Chem.* **2016**, *8*, 803−16.
(31) Lorenz, D. A.; Song, J. M.; Garner, A. L. High-throughput platform assay technology for the discovery of pre-microrna-selective small molecule probes. *Bioconjugate Chem.* **2015**, *26*, 19−23.
(32) Disney, M. D.; Winkelsas, A. M.; Velagapudi, S. P.; Southern, M.; Fallahi, M.; Childs-Disney, J. L. Inforna 2.0: A Platform for the Sequence-Based Design of Small Molecules Targeting Structured RNAs. *ACS Chem. Biol.* **2016**, *11*, 1720−8.
(33) Haga, C. L.; Velagapudi, S. P.; Childs-Disney, J. L.; Strivelli, J.; Disney, M. D.; Phinney, D. G. Rapid Generation of miRNA Inhibitor Leads by Bioinformatics and Efficient High-Throughput Screening Methods. *Methods Mol. Biol. (N. Y., NY, U. S.)* **2017**, *1517*, 179−198.
(34) Chen, X.; Yan, C. C.; Zhang, X.; Zhang, X.; Dai, F.; Yin, J.; Zhang, Y. Drug-target interaction prediction: databases, web servers and computational models. *Briefings Bioinf.* **2016**, *17*, 696−712.
(35) Chen, X.; Guan, N. N.; Sun, Y. Z.; Li, J. Q.; Qu, J. MicroRNA-small molecule association identification: from experimental results to computational models. *Briefings Bioinf.* **2018**, DOI: 10.1093/bib/bby098.
(36) Jiang, W.; Chen, X.; Liao, M.; Li, W.; Lian, B.; Wang, L.; Meng, F.; Liu, X.; Chen, X.; Jin, Y.; Li, X. Identification of links between small molecules and miRNAs in human cancers based on transcriptional responses. *Sci. Rep.* **2012**, *2*, 282.
(37) Meng, F.; Dai, E.; Yu, X.; Zhang, Y.; Chen, X.; Liu, X.; Wang, S.; Wang, L.; Jiang, W. Constructing and characterizing a bioactive small molecule and microRNA association network for Alzheimer's disease. *J. R. Soc., Interface* **2014**, *11*, 20131057.
(38) Lv, Y.; Wang, S.; Meng, F.; Yang, L.; Wang, Z.; Wang, J.; Chen, X.; Jiang, W.; Li, Y.; Li, X. Identifying novel associations between small molecules and miRNAs based on integrated molecular networks. *Bioinformatics* **2015**, *31*, 3638−44.
(39) Li, J.; Lei, K.; Wu, Z.; Li, W.; Liu, G.; Liu, J.; Cheng, F.; Tang, Y. Network-based identification of microRNAs as potential pharmacogenomic biomarkers for anticancer drugs. *Oncotarget* **2016**, *7*, 45584−45596.
(40) Wang, J.; Meng, F.; Dai, E.; Yang, F.; Wang, S.; Chen, X.; Yang, L.; Wang, Y.; Jiang, W. Identification of associations between small molecule drugs and miRNAs based on functional similarity. *Oncotarget* **2016**, *7*, 38658−38669.
(41) Qu, J.; Chen, X.; Sun, Y. Z.; Li, J. Q.; Ming, Z. Inferring potential small molecule-miRNA association based on triple layer heterogeneous network. *J. Cheminf.* **2018**, *10*, 30.
(42) Guan, N. N.; Sun, Y. Z.; Ming, Z.; Li, J. Q.; Chen, X. Prediction of Potential Small Molecule-Associated MicroRNAs Using Graphlet Interaction. *Front. Pharmacol.* **2018**, *9*, 1152.
(43) Xu, T.; Li, B.; Zhao, M.; Szulwach, K. E.; Street, R. C.; Lin, L.; Yao, B.; Zhang, F.; Jin, P.; Wu, H.; Qin, Z. S. Base-resolution methylation patterns accurately predict transcription factor bindings in vivo. *Nucleic Acids Res.* **2015**, *43*, 2757−66.
(44) Rajagopal, N.; Xie, W.; Li, Y.; Wagner, U.; Wang, W.; Stamatoyannopoulos, J.; Ernst, J.; Kellis, M.; Ren, B. RFECS: a random-forest based algorithm for enhancer identification from chromatin state. *PLoS Comput. Biol.* **2013**, *9*, e1002968.
(45) Ulintz, P. J.; Zhu, J.; Qin, Z. S.; Andrews, P. C. Improved classification of mass spectrometry database search results using newer machine learning approaches. *Mol. Cell. Proteomics* **2006**, *5*, 497−509.
(46) Chen, X.; Wang, C. C.; Yin, J.; You, Z. H. Novel Human miRNA-Disease Association Inference Based on Random Forest. *Mol. Ther.–Nucleic Acids* **2018**, *13*, 568−579.
(47) Ritchie, G. R.; Dunham, I.; Zeggini, E.; Flicek, P. Functional annotation of noncoding sequence variants. *Nat. Methods* **2014**, *11*, 294−6.
(48) Chen, L.; Jin, P.; Qin, Z. S. DIVAN: accurate identification of non-coding disease-specific risk variants using multi-omics profiles. *Genome Biol.* **2016**, *17*, 252.
(49) Liu, X.; Wang, S.; Meng, F.; Wang, J.; Zhang, Y.; Dai, E.; Yu, X.; Li, X.; Jiang, W. SM2miR: a database of the experimentally validated small molecules' effects on microRNA expression. *Bioinformatics* **2013**, *29*, 409−11.
(50) Knox, C.; Law, V.; Jewison, T.; Liu, P.; Ly, S.; Frolkis, A.; Pon, A.; Banco, K.; Mak, C.; Neveu, V.; Djoumbou, Y.; Eisner, R.; Guo, A. C.; Wishart, D. S. DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res.* **2011**, *39*, D1035−41.
(51) Wang, Y.; Xiao, J.; Suzek, T. O.; Zhang, J.; Wang, J.; Bryant, S. H. PubChem: a public information system for analyzing bioactivities of small molecules. *Nucleic Acids Res.* **2009**, *37*, W623−33.
(52) 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*, D98−104.
(53) Lu, M.; Zhang, Q.; Deng, M.; Miao, J.; Guo, Y.; Gao, W.; Cui, Q. An analysis of human microRNA and disease associations. *PLoS One* **2008**, *3*, e3420.

DOI: 10.1021/acs.jcim.9b00129
J. Chem. Inf. Model. 2019, 59, 1668−1679

(54) Ruepp, A.; Kowarsch, A.; Schmidl, D.; Bruggenthin, F.; Brauner, B.; Dunger, I.; Fobo, G.; Frishman, G.; Montrone, C.; Theis, F. J. PhenomiR: a knowledgebase for microRNA expression in diseases and biological processes. *Genome Biol.* **2010**, *11*, R6.
(55) Gottlieb, A.; Stein, G. Y.; Ruppin, E.; Sharan, R. PREDICT: a method for inferring novel drug indications with application to personalized medicine. *Mol. Syst. Biol.* **2011**, *7*, 496.
(56) Lv, S.; Li, Y.; Wang, Q.; Ning, S.; Huang, T.; Wang, P.; Sun, J.; Zheng, Y.; Liu, W.; Ai, J.; Li, X. A novel method to quantify gene set functional association based on gene ontology. *J. R. Soc., Interface* **2012**, *9*, 1063−72.
(57) Hattori, M.; Okuno, Y.; Goto, S.; Kanehisa, M. Development of a chemical structure comparison method for integrated analysis of chemical and genomic information in the metabolic pathways. *J. Am. Chem. Soc.* **2003**, *125*, 11853−65.
(58) Cheng, Z.; Huang, K.; Wang, Y.; Liu, H.; Guan, J.; Zhou, S. Selecting high-quality negative samples for effectively predicting protein-RNA interactions. *BMC Syst. Biol.* **2017**, *11*, 9.
(59) Peng, L.; Zhu, W.; Liao, B.; Duan, Y.; Chen, M.; Chen, Y.; Yang, J. Screening drug-target interactions with positive-unlabeled learning. *Sci. Rep.* **2017**, *7*, 8087.
(60) Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5−32.
(61) Heidelberger, C.; Chaudhuri, N. K.; Danneberg, P.; Mooren, D.; Griesbach, L.; Duschinsky, R.; Schnitzer, R. J.; Pleven, E.; Scheiner, J. Fluorinated pyrimidines, a new class of tumour-inhibitory compounds. *Nature* **1957**, *179*, 663−6.
(62) Longley, D. B.; Harkin, D. P.; Johnston, P. G. 5-fluorouracil: mechanisms of action and clinical strategies. *Nat. Rev. Cancer* **2003**, *3*, 330−8.
(63) Hess, A. K.; Muer, A.; Mairinger, F. D.; Weichert, W.; Stenzinger, A.; Hummel, M.; Budach, V.; Tinhofer, I. MiR-200b and miR-155 as predictive biomarkers for the efficacy of chemoradiation in locally advanced head and neck squamous cell carcinoma. *Eur. J. Cancer* **2017**, *77*, 3−12.
(64) Khorrami, S.; Zavaran Hosseini, A.; Mowla, S. J.; Soleimani, M.; Rakhshani, N.; Malekzadeh, R. MicroRNA-146a induces immune suppression and drug-resistant colorectal cancer cells. *Tumor Biol.* **2017**, *39*, 101042831769836.
(65) Shen, W. W.; Zeng, Z.; Zhu, W. X.; Fu, G. H. MiR-142−3p functions as a tumor suppressor by targeting CD133, ABCG2, and Lgr5 in colon cancer cells. *J. Mol. Med. (Heidelberg, Ger.)* **2013**, *91*, 989−1000.
(66) Song, B.; Wang, Y.; Xi, Y.; Kudo, K.; Bruheim, S.; Botchkina, G. I.; Gavin, E.; Wan, Y.; Formentini, A.; Kornmann, M.; Fodstad, O.; Ju, J. Mechanism of chemoresistance mediated by miR-140 in human osteosarcoma and colon cancer cells. *Oncogene* **2009**, *28*, 4065−74.
(67) Liu, H.; Yin, Y.; Hu, Y.; Feng, Y.; Bian, Z.; Yao, S.; Li, M.; You, Q.; Huang, Z. miR-139−5p sensitizes colorectal cancer cells to 5-fluorouracil by targeting NOTCH-1. *Pathol., Res. Pract.* **2016**, *212*, 643−9.
(68) Ascenzi, P.; Bocedi, A.; Marino, M. Structure-function relationship of estrogen receptor alpha and beta: impact on human health. *Mol. Aspects Med.* **2006**, *27*, 299−402.
(69) Jick, S. S.; Hagberg, K. W.; Kaye, J. A.; Jick, H. Postmenopausal estrogen-containing hormone therapy and the risk of breast cancer. *Obstet. Gynecol.* **2009**, *113*, 74−80.
(70) Jacovetti, C.; Abderrahmani, A.; Parnaud, G.; Jonas, J. C.; Peyot, M. L.; Cornu, M.; Laybutt, R.; Meugnier, E.; Rome, S.; Thorens, B.; Prentki, M.; Bosco, D.; Regazzi, R. MicroRNAs contribute to compensatory beta cell expansion during pregnancy and obesity. *J. Clin. Invest.* **2012**, *122*, 3541−51.
(71) Tao, S.; He, H.; Chen, Q. Estradiol induces HOTAIR levels via GPER-mediated miR-148a inhibition in breast cancer. *J. Transl. Med.* **2015**, *13*, 131.
(72) Jeyabalan, J.; Aqil, F.; Munagala, R.; Annamalai, L.; Vadhanam, M. V.; Gupta, R. C. Chemopreventive and therapeutic activity of dietary blueberry against estrogen-mediated breast cancer. *J. Agric. Food Chem.* **2014**, *62*, 3963−71.
(73) Zierau, O.; Helle, J.; Schadyew, S.; Morgenroth, Y.; Bentler, M.; Hennig, A.; Chittur, S.; Tenniswood, M.; Kretzschmar, G. Role of miR-203 in estrogen receptor-mediated signaling in the rat uterus and endometrial carcinoma. *J. Cell. Biochem.* **2018**, *119*, 5359−5372.
(74) Momparler, R. L. Pharmacology of 5-Aza-2′-deoxycytidine (decitabine). *Semin. Hematol.* **2005**, *42*, S9−16.
(75) Momparler, R. L.; Rivard, G. E.; Gyger, M. Clinical trial on 5-aza-2′-deoxycytidine in patients with acute leukemia. *Pharmacol. Ther.* **1985**, *30*, 277−86.
(76) Malik, P.; Cashen, A. F. Decitabine in the treatment of acute myeloid leukemia in elderly patients. *Cancer Manage. Res.* **2014**, *6*, 53−61.
(77) Lee, Y. M.; Chen, H. W.; Maurya, P. K.; Su, C. M.; Tzeng, C. R. MicroRNA regulation via DNA methylation during the morula to blastocyst transition in mice. *Mol. Hum. Reprod.* **2012**, *18*, 184−93.
(78) Xu, W. Q.; Huang, Y. M.; Xiao, H. F. Expression Analysis and Epigenetics of MicroRNA let-7b in Acute Lymphoblastic Leukemia. *Zhongguo Shiyan Xueyexue Zazhi* **2015**, *23*, 1535−41.
(79) Deng, H.; Guo, Y.; Song, H.; Xiao, B.; Sun, W.; Liu, Z.; Yu, X.; Xia, T.; Cui, L.; Guo, J. MicroRNA-195 and microRNA-378 mediate tumor growth suppression by epigenetical regulation in gastric cancer. *Gene* **2013**, *518*, 351−9.
(80) Zhang, Y.; Cheng, C.; He, D.; Shi, W.; Fu, C.; Wang, X.; Zeng, C. Transcriptional gene silencing of dopamine D3 receptor caused by let-7d mimics in immortalized renal proximal tubule cells of rats. *Gene* **2016**, *580*, 89−95.
(81) Chen, X.; Wang, L.; Qu, J.; Guan, N. N.; Li, J. Q. Predicting miRNA-disease association based on inductive matrix completion. *Bioinformatics* **2018**, *34*, 4256−4265.
(82) Chen, X.; Xie, D.; Wang, L.; Zhao, Q.; You, Z. H.; Liu, H. BNPMDA: Bipartite Network Projection for MiRNA-Disease Association prediction. *Bioinformatics* **2018**, *34*, 3178−3186.
(83) Chen, X.; Yin, J.; Qu, J.; Huang, L. MDHGI: Matrix Decomposition and Heterogeneous Graph Inference for miRNA-disease association prediction. *PLoS Comput. Biol.* **2018**, *14*, e1006418.
(84) Chen, X.; Huang, L.; Xie, D.; Zhao, Q. EGBMMDA: Extreme Gradient Boosting Machine for MiRNA-Disease Association prediction. *Cell Death Dis.* **2018**, *9*, 3.
(85) Huang, Y. A.; Chan, K. C. C.; You, Z. H. Constructing prediction models from expression profiles for large scale lncRNA-miRNA interaction profiling. *Bioinformatics* **2018**, *34*, 812−819.
(86) Chen, X.; Ren, B.; Chen, M.; Wang, Q.; Zhang, L.; Yan, G. NLLSS: Predicting Synergistic Drug Combinations Based on Semi-supervised Learning. *PLoS Comput. Biol.* **2016**, *12*, e1004975.
(87) Chen, X.; Sun, Y. Z.; Guan, N. N.; Qu, J.; Huang, Z. A.; Zhu, Z. X.; Li, J. Q. Computational models for lncRNA function prediction and functional similarity calculation. *Briefings Funct. Genomics* **2019**, *18*, 58.
(88) Chen, X.; Yan, C. C.; Zhang, X.; You, Z. H. Long non-coding RNAs and complex diseases: from experimental results to computational models. *Briefings Bioinf.* **2016**, *18*, 558−576.
(89) Chen, X.; Sun, Y. Z.; Zhang, D. H.; Li, J. Q.; Yan, G. Y.; An, J. Y.; You, Z. H. NRDTD: a database for clinically or experimentally supported non-coding RNAs and drug targets associations. *Database* **2017**, *2017*, 2017.
(90) Wang, E.; Zaman, N.; McGee, S.; Milanese, J. S.; Masoudi-Nejad, A.; O'Connor-McCourt, M. Predictive genomics: a cancer hallmark network framework for predicting tumor clinical phenotypes using genome sequencing data. *Semin. Cancer Biol.* **2015**, *30*, 4−12.

DOI: 10.1021/acs.jcim.9b00129
J. Chem. Inf. Model. 2019, 59, 1668−1679