Hindawi
Complexity
Volume 2017, Article ID 2713280, 10 pages
https://doi.org/10.1155/2017/2713280

Research Article

SDTRLS: Predicting Drug-Target Interactions for Complex Diseases Based on Chemical Substructures

Cheng Yan,[1,2] Jianxin Wang,[1] Wei Lan,[1] Fang-Xiang Wu,[3] and Yi Pan[4]

[1]*School of Information Science and Engineering, Central South University, Changsha, Hunan 410083, China*
[2]*School of Computer and Information, Qiannan Normal University for Nationalities, Duyun, Guizhou 558000, China*
[3]*Department of Mechanical Engineering and Division of Biomedical Engineering, University of Saskatchewan, Saskatoon, SK, Canada*
[4]*Department of Computer Science, Georgia State University, Atlanta, GA 30302, USA*

Correspondence should be addressed to Jianxin Wang; jxwang@mail.csu.edu.cn

Received 8 April 2017; Revised 19 October 2017; Accepted 1 November 2017; Published 3 December 2017

Academic Editor: Daniela Paolotti

It is well known that drug discovery for complex diseases via biological experiments is a time-consuming and expensive process. Alternatively, the computational methods provide a low-cost and high-efficiency way for predicting drug-target interactions (DTIs) from biomolecular networks. However, the current computational methods mainly deal with DTI predictions of known drugs; there are few methods for large-scale prediction of failed drugs and new chemical entities that are currently stored in some biological databases may be effective for other diseases compared with their originally targeted diseases. In this study, we propose a method (called SDTRLS) which predicts DTIs through RLS-Kron model with chemical substructure similarity fusion and Gaussian Interaction Profile (GIP) kernels. SDTRLS can be an effective predictor for targets of old drugs, failed drugs, and new chemical entities from large-scale biomolecular network databases. Our computational experiments show that SDTRLS outperforms the state-of-the-art SDTNBI method; specifically, in the G protein-coupled receptors (GPCRs) external validation, the maximum and the average AUC values of SDTRLS are 0.842 and 0.826, respectively, which are superior to those of SDTNBI, which are 0.797 and 0.766, respectively. This study provides an important basis for new drug development and drug repositioning based on biomolecular networks.

1. Introduction

The identification of target molecules associated with specific diseases is the basis of modern drug discovery and development [1–3]. Therefore, the identification of drug-target interactions (DTIs) is important for drug development. However, it is well known that drug discovering is a cost- and time-consuming process in the field of pharmacology. According to the USA Food and Drug Administration statistical data, the cost of new drug discovery is approximately $1.8 billion and it takes an average of 13 years [4]. Therefore, how to deal with this problem becomes an emerging issue. Over decades, different computational methods and tools [5–13] have been developed to predict large-scale potential DTIs and drug repositing through the unremitting efforts of a large number of researchers and organizations under the development of computing technology.

Meanwhile, many DTI data have been generated with the rapid growth of the public chemical and biological database. For example, PubChem [14] is a freely available chemistry database. There are 7759 drug entities, 4104 target proteins, and 15,199 DTIs present in DrugBank [15] database by now. The freely available online ChEMBL [16] database provides pharmaceutical chemists with a convenient platform for querying target bioactivity data for compounds or targets. In addition, including TTD [17], KEGG [18], SIDER [19], STITCH [20], STRING [21], BindingDB [22], and other various kinds of resources have established the basis for DTI prediction.

Now it is possible for us to quickly and inexpensively identify potential DTIs and repurpose existing drugs [23–27] through the developments of computational methods. These methods are mainly divided into three categories, including basic network-based models, machine learning-based models, and other approaches based on similarity [28].

From the viewpoint of basic network-based model, Cheng et al. [29] developed the method to predict DTIs through network-based inference (NBI). Comparing with drug-based similarity inference (DBSI) and target-based similarity inference (TBSI), NBI is better than them because it is in the full use of the known DTIs. Moreover, node- and edge-weighted NBI was developed via constructing the weight of nodes and edges on drug-target network. Network-based Random Walk with Restart on the Heterogeneous network (NRWRH) was developed by Chen et al., which implemented the random walk on the heterogeneous network (protein-protein similarity network, drug-drug similarity network, and known drug-target interaction networks) [30]. It is an enhanced version of the traditional random walk that improved the predictive performance through making full use of data with the integrated heterogeneous network.

Some machine learning-based approaches were also developed to predict DTIs. Following Bleakley et al. [31] and Mordelet and Vert [32], Bleakley and Yamanishi [33] further proposed the bipartite local model (BLM) to predict DTIs, which used local support vector machine (SVM) classifiers with known DTIs and integrated the chemical structure similarity and protein sequence similarity information. Gaussian Interaction Profile (GIP) kernels on drug-target networks were significant improvements developed by van Laarhoven et al. [34]. In order to solve the problem of negative samples, Lan et al. [35] proposed a prediction method (PUDT) which classified unlabeled samples into the reliable negative examples and likely negative examples based on the similarity of protein structure and achieved good results.

The matrix decomposition technique is also used for predicting DTIs, miRNA-disease associations [36, 37], and so on. It maps the DTI matrix to the low-dimensional matrix to infer the hidden interactions based on the known interactions. Gönen [38] proposed a Bayesian model that combined dimensionality reduction, matrix factorization, and binary classification for predicting DTIs via integrating the drug-drug chemical similarity and protein-protein sequence similarity. Multiple Similarities Collaborative Matrix Factorization (MSCMF) [39] method projected drugs and targets into a common low-rank feature space and significantly improved the results via adjusting the weight of similarity matrix of drugs and of targets. Ezzat et al. [40] developed the Regularized Matrix Factorization method that distinguished from many of nonoccurring edges in the interaction matrix which are actually unknown or hide cases by other similarity information. DrugE-Rank [12] developed a machine learning-based model by combining the advantages of two different types of feature-based and similarity-based methods to improve the prediction performance.

Although the above methods have gained good results in predicting the new DTIs on known drugs, it is also important to predict DTIs of failed drugs and new chemical entities. There are thousands of drugs that are failed in clinical phases and even US National Center for Advancing Translational Sciences is paying US$20 million to research for repurposing 58 failed drugs [41, 42] as the drugs that failed in their initially targeted diseases may be effective in other diseases. Wu et al. [43] proposed an integrated network and chemoinformatics tool for systematic prediction of DTIs and drug repositioning, namely, SDTNBI (substructure-drug-target network-based inference) which predicted new DTIS of failed drugs and new chemical entities by integrating known DTIS and chemical substructure of failed drugs or new chemical entities in a way of resource diffusion. Their study assumed that chemical substructure played key roles in DTIs. This method achieved good prediction results for large-scale failed drugs and new chemical entities based on chemical substructures shared between them and the known drugs.

In this study, we propose a method called SDTRLS (substructure-drug-target Kronecker product kernel regularized least squares) for large-scale DTI prediction and drug repositioning based on the chemical substructures of known drugs, failed drugs, and new chemical entities. Firstly, we compute the substructure similarity and then create a Gaussian Interaction Profile (GIP) kernels for drug entities and target proteins based on known DTIs. The k-nearest neighbor (KNN) was used to compute the initial relational score in the presence of a new chemical entity or failed drug that has no known DTIs. Through similarity network fusion (SNF) technology [44] the similarity of substructure and GIP of drugs are integrated. SNF substantially outperforms single-type data analysis and establishes integrative approaches to predicting DTIs. Finally, the RLS-Kron [34] classifier was used to predict DTIs, which constructs a large kernel that directly relates to the drug-target pairs by combining the similarity kernels of drug entities and target proteins. In order to comprehensively assess the performance of our method, we compare it against current state-of-the-art algorithms with the same data and evaluation criteria. We use the 10-fold cross validation and external validation to show the accuracy and robustness of our method. The computational results show that our proposed SDTRLS is comparable to other five methods in terms of stability. Especially in the G protein-coupled receptors (GPCRs) external validation dataset, the maximum and average AUC values were 0.842 and 0.826, respectively, which are superior to 0.797 and 0.766 from state-of-the-art SDTNBI method. In order to further confirm the prediction ability of STDRLS, we perform an experimental analysis on some prediction results. In summary, we provide a new alternative method for DTI prediction for known drugs, failed drugs, and new chemical entities. It provides the basis for drug discovery, development, and personalized medical treatment in the future.

2. Materials

This study used five internal validation datasets and two external validation datasets. The internal datasets are used to validate the predictions of the new DTIs of known drugs, and the external datasets are used to validate the predictions of all DTIs of new entities and failed drugs. Five internal datasets are G protein-coupled receptors (GPCRs), kinase superfamily (Kinases), ion channels (ICs), nuclear receptors (NRs), and Global. GPCRs and Kinases were downloaded from ChEMBL database. ICs and NRs were collected from the ChEMBL and BindingDB database. The Global is a global network covering genomewide targets where all drugs also

Table 1: Drugs, targets, and DTIs in each dataset.

Datasets	Targets	S_d	S_t	N_{dt}	Sparsity (%)
Internal datasets	GPCRs	4741	97	17,111	3.72
	Kinases	2827	206	13,647	2.34
	ICs	7929	97	8944	1.16
	NRs	5218	35	7366	4.03
	Global	1844	1032	10,185	0.54
External datasets	ExGPCRs	92	46	271	6.4
	ExKinases	188	28	202	3.84

S_d is the number of drugs, S_t is the number of targets, N_{dt} is the known DTIs, and sparsity is the proportion of the of N_{dt} to all possible DTIs in datasets.

come from DrugBank database. Two external datasets were selected from GPCRs and Kinases in DrugBank database, respectively.

The external validation is to predict all DTIs for drugs, so it needs a basic dataset that includes drugs, targets, and known DTIs. GPCRs and Kinases are the basic datasets to ExGPCRs and ExKinases, respectively. The known 17,111 DTIs of GPCRs are the prior knowledge to external validation of ExGPCRs in Table 1.

Table 1 shows that the 92 drugs of ExGPCRs and 4741 of GPCRs are independent of each other. However, the 46 targets of ExGPCRs are the subset of the 92 targets of GPCRs. Furthermore, the relationship of drugs and targets between Kinases and ExKinases is the same as that between ExGPCRs and GPCRs. These datasets can be downloaded from http://lmmd.ecust.edu.cn/methods/sdtnbi/#*. Table 1 contains some statistics of five internal validation datasets and two external datasets.

2.1. Chemical Substructure. In this study, we used seven types of fingerprints to express the chemical substructures of each molecule. All substructure data are generated from PaDEL-Descriptor software, including CDK Fingerprint, CDK Extended Fingerprint, CDK Graph Only Fingerprint, Substructure Fingerprint, Klekota-Roth Fingerprint, MACCS Fingerprint, and PubChem Fingerprint, namely, CDK, CDKExt, Graph, FP4, KR, MACCS, and PubChem, respectively. Each type of substructures of each molecule is represented by a multiple dimensional vector with values of 0 or 1. We only used the substructures that appear in the datasets.

Table 2 contains the overview of the seven substructures of dataset GPCRs, including the dimension of each chemical substructure. The dimensions of substructures were derived from the statistics result of the datasets that include all appearing substructure types.

Table 2: The dimensions on GPCRs.

Chemical substructure types	Dimensions
CDK	1024
CDKExt	1012
FP4	131
Graph	1023
KR	1834
MACCS	153
PubChem	627

3. Methods

3.1. Chemical Substructure Similarity. Let $S = \{s_1, s_2, \ldots, s_K\}$ be a set of all substructures for one type of seven chemical substructures, where K is the dimension of the chemical substructure. For example, the value of K is 1024 in CDK and the value of K is 153 in MACCS. $D = \{d_1, d_2, \ldots, d_m\}$ is the set of drugs, where m is the number of drugs. For one chemical substructure, drug d_i can be represented by a profile (binary vector) of the substructure, that is, $DS(d_i) = \{ds_1(d_i), ds_2(d_i), \ldots, ds_K(d_i)\}$. If drug d_i has s_k, the value of $ds_k(d_i)$ is 1, otherwise 0. For a type of chemical substructure, the substructure similarity $S_{\text{subsim}}(d_i, d_j)$ of drugs d_i and d_j can be computed by the weighted cosine correlation coefficient based on the substructure information [27].

$$S_{\text{subsim}}\left(d_i, d_j\right) = \frac{\sum_{k=1}^{K} w_k ds_k\left(d_i\right) ds_k\left(d_j\right)}{\sqrt{\sum_{k=1}^{K} w_k ds_k^2\left(d_i\right)} \sqrt{\sum_{k=1}^{K} w_k ds_k^2\left(d_j\right)}}, \quad (1)$$

where w_k is the weight of the kth substructure (s_k), which can be calculated by the formula [27]

$$w_k = \exp\left(-\frac{f_k^2}{\delta^2 h^2}\right), \quad (2)$$

where f_k is the frequency of chemical substructure s_k in the whole dataset, δ is the standard deviation of $\{s_k\}_{k=1}^{k=K}$, and h is a parameter (set to be 0.1 in this study). The basic rationale for introducing the weight to compute substructure similarity between drugs and new chemical entities is that substructures with fewer occurrences should occupy a more proportion than substructures which appear frequently.

3.2. Gaussian Interaction Profile Kernel. We denoted that $T = \{t_1, t_2, \ldots, t_n\}$ is the set of n targets. A drug-target network can be represented by a bipartite graph which has an adjacency

matrix $Y \in R^{m*n}$, where the value of y_{ij} is 1 if d_i and t_j have known DTI, otherwise 0. The Gaussian Interaction Profile (GIP) kernel is constructed from the topology information of known DTIs network [10, 34]. The kernel of drugs d_i and d_j can be formulated as

$$K_{\mathrm{GIP},d}\left(d_i, d_j\right) = \exp\left(-\gamma_d \left\|Y\left(d_i\right) - Y\left(d_j\right)\right\|^2\right),$$
$$\gamma_d = \frac{\alpha}{\left(\left(1/N_d\right)\sum_{i=1}^{N_d}\left\|Y\left(d_i\right)\right\|^2\right)}, \tag{3}$$

where $Y(d_i) = \{y_{i1}, y_{i2}, \ldots, y_{in}\}$ is the interaction profile of drug d_i, and α is a parameter that controls the bandwidth; we set the value to be 1 in this study. Similarly, the kernel of targets t_i and t_j can be calculated by (4).

$$K_{\mathrm{GIP},t}\left(t_i, t_j\right) = \exp\left(-\gamma_t \left\|Y\left(t_i\right) - Y\left(t_j\right)\right\|^2\right), \tag{4}$$

$$\gamma_t = \frac{\beta}{\left(\left(1/N_t\right)\sum_{i=1}^{N_t}\left\|Y\left(t_i\right)\right\|^2\right)}, \tag{5}$$

where $Y(t_j) = \{y_{1j}, y_{2j}, \ldots, y_{mj}\}^T$ is the interaction profile of target t_j; we also set the parameter β to be 1.

3.3. Similarity Network Fusion. We have two similarity matrices for drugs (including known drugs, new chemical entities), namely, substructure similarity $S_{\mathrm{subsim}} \in R^{m*m}$ and $K_{\mathrm{GIP},d} \in R^{m*m}$. To construct more comprehensive similarity kernel for drugs, we used the SNF method to fuse two similarity kernels.

Firstly, the row-normalized matrices $P^{(1)}$ and $P^{(2)}$ are calculated from the drug similarity matrices S_{subsim} and $K_{\mathrm{GIP},d}$, respectively. Secondly, according to the K-nearest neighbors (KNN) method, the resultant matrices $S^{(1)}$ and $S^{(2)}$ are obtained from $P^{(1)}$ and $P^{(2)}$ by the following equation[44]:

$$S\left(d_i, d_j\right) = \begin{cases} \dfrac{P\left(d_i, d_j\right)}{\sum_{d_k \in N(d_i)} P\left(d_i, d_k\right)}, & d_j \in N\left(d_i\right), \\ 0, & \text{otherwise}, \end{cases} \tag{6}$$

where $N(d_i)$ is the set of top N similar neighbors of drug d_i. In this study, we set the value of N to be 50. The main idea of SNF is iteratively updating similarity matrices $P^{(1)}$ and $P^{(2)}$ [44].

$$P_{t+1}^{(1)} = S^{(1)} \times P_t^{(2)} \times \left(S^{(1)}\right)^T,$$
$$P_{t+1}^{(2)} = S^{(2)} \times P_t^{(1)} \times \left(S^{(2)}\right)^T, \tag{7}$$

where the parameter t represents the times of iterations, and its value is set to be 20 in this study by considering that the iteration time can not be too long and $\max(\max(\mathrm{abs}(((P_t^{(1)} + P_t^{(2)})/2 - (P_{t-1}^{(1)} + P_{t-1}^{(2)})/2)))) < 10^{-3}$. The initial matrices are defined as $P_{t=1}^{(1)} = P^{(1)}$ and $P_{t=1}^{(2)} = P^{(2)}$. The final similarity matrix $S_{\mathrm{final}} \in R^{m*m}$ of drugs is calculated from the average value of matrices $P_{20}^{(1)}$ and $P_{20}^{(2)}$ ($S_{\mathrm{final}} = (P_{20}^{(1)} + P_{20}^{(2)})/2$).

3.4. Kron_RLS. Kronecker product kernels are used widely in prediction issues of other studies and conditions [45–47]. In this study, we also use a Kronecker product kernel to construct a larger kernel for the drug-target pairs. Then the prediction of DTIs is based on the ranking of the pairs that include known drugs and targets and new entities or failed drugs and targets. The higher rank implies the higher possibility of existing interactions. Based on the kernel of drugs and targets, the Kronecker product kernel of drug-target pairs is constructed as follows [34]:

$$K\left(\left(d_i, t_j\right), \left(d_k, t_l\right)\right) = K_d\left(d_i, d_k\right) K_t\left(t_j, t_l\right), \tag{8}$$

where $K_d(d_i, d_k)$ is the (i,k)th element of the kernel of drugs with S_{final}, while $K_t(t_j, t_l)$ is the (j,l)th element kernel of targets with $K_{\mathrm{GIP},t}$.

According the Kronecker product kernel of formula (8), the predictions of DTIs for all drug-target pairs can be calculated as follows [34]:

$$\mathrm{vec}\left(\widehat{Y}^T\right) = K\left(K + \sigma I\right)^{-1} \mathrm{vec}\left(Y^T\right), \tag{9}$$

where σ is a regularization parameter. The smoother result can be obtained via the higher value σ. We get $\widehat{Y} = Y$ when $\sigma = 0$ which shows no generalization [34]. We also use the eigendecompositions of the kernel matrices according to Laarhoven's study. The eigendecompositions of matrices K_d and K_t are $K_d = \vee_d \wedge_d \vee_d^T$ and $K_t = \vee_t \wedge_t \vee_t^T$, in which $\vee_d$ and $\vee_t$ are the unitary matrices of feature vectors, and $\wedge_d$ and $\wedge_t$ are the diagonal matrices of eigenvalues for drugs and targets, respectively. Since the eigenvalues (vectors) of a Kronecker product are the Kronecker product of eigenvalues (vectors), the Kronecker product kernel of drug-target pairs can be formulated as follows [34]:

$$K = K_d \otimes K_t = \vee \wedge \vee^T, \tag{10}$$

in which

$$\vee = \vee_d \otimes \vee_t,$$
$$\wedge = \wedge_d \otimes \wedge_t. \tag{11}$$

3.5. KNN for New Chemical Entities. New chemical entities or failed drugs have no known associations with targets, which makes it impossible to predict more associations by existing methods. In this study, we used the KNN method to estimate the interaction scores for new chemical entities or failed drugs by the similarity between them and known drugs. For example, we denote a new chemical entity or failed drug as C_{new}, whose interaction score with target t_j can be computed by the formula

$$\mathrm{Score}\left(C_{\mathrm{new}}, t_j\right) = \frac{\sum S_{\mathrm{subsim}}^{(d_i,d_l)} y_{lj}}{\sum S_{\mathrm{subsim}}^{(d_i,d_l)}}, \quad l \in K_{\mathrm{new}}, \tag{12}$$

where $S_{\mathrm{subsim}}^{(d_i,d_l)}$ is the (i,l)th element of chemical substructure similarity matrix $S_{\mathrm{subsim}} \in R^{m*m}$, and y_{lj} is the (l,j)th element of $Y \in R^{m*n}$. K_{new} is the set of top K neighbors according to the S_{subsim} matrix. In this study, we set the value of K to be 4.

Table 3: The performance of 10-fold cross validation on 5 datasets.

		AUC					
Target	FP	DBSI-R	NWNBI	EWNBI*	NBI	SDTNBI*	SDTRLS
GPCRs	CDK	0.896 ± 0.003	0.981 ± 0.001	0.981 ± 0.001	0.980 ± 0.001	0.904 ± 0.003	0.982 ± 0.002
	CDKExt	0.895 ± 0.002	0.981 ± 0.001	0.981 ± 0.001	0.980 ± 0.001	0.901 ± 0.003	0.982 ± 0.002
	FP4	0.896 ± 0.002	0.981 ± 0.001	0.981 ± 0.001	0.980 ± 0.001	0.966 ± 0.002	0.979 ± 0.002
	Graph	0.897 ± 0.002	0.981 ± 0.001	0.981 ± 0.001	0.980 ± 0.001	0.917 ± 0.003	0.980 ± 0.001
	KR	0.909 ± 0.002	0.981 ± 0.001	0.981 ± 0.001	0.980 ± 0.001	0.960 ± 0.002	0.983 ± 0.002
	MACCS	0.881 ± 0.005	0.981 ± 0.001	0.981 ± 0.001	0.980 ± 0.001	0.931 ± 0.002	0.982 ± 0.001
	PubChem	0.895 ± 0.003	0.981 ± 0.001	0.981 ± 0.001	0.980 ± 0.001	0.918 ± 0.003	0.981 ± 0.001
Kinases	CDK	0.886 ± 0.004	0.977 ± 0.002	0.976 ± 0.002	0.976 ± 0.002	0.893 ± 0.004	0.972 ± 0.002
	CDKExt	0.885 ± 0.002	0.977 ± 0.002	0.976 ± 0.002	0.976 ± 0.002	0.892 ± 0.004	0.973 ± 0.002
	FP4	0.880 ± 0.004	0.977 ± 0.002	0.976 ± 0.002	0.976 ± 0.002	0.956 ± 0.003	0.969 ± 0.003
	Graph	0.882 ± 0.003	0.977 ± 0.002	0.976 ± 0.002	0.976 ± 0.002	0.905 ± 0.003	0.971 ± 0.003
	KR	0.901 ± 0.003	0.977 ± 0.002	0.976 ± 0.002	0.976 ± 0.002	0.958 ± 0.003	0.970 ± 0.003
	MACCS	0.880 ± 0.002	0.977 ± 0.002	0.976 ± 0.002	0.976 ± 0.002	0.925 ± 0.003	0.970 ± 0.003
	PubChem	0.877 ± 0.005	0.977 ± 0.002	0.976 ± 0.002	0.976 ± 0.002	0.903 ± 0.003	0.971 ± 0.002
ICs	CDK	0.923 ± 0.002	0.582 ± 0.007	Ø	0.573 ± 0.013	0.932 ± 0.004	0.956 ± 0.005
	CDKExt	0.922 ± 0.003	0.582 ± 0.007	Ø	0.573 ± 0.013	0.931 ± 0.004	0.955 ± 0.005
	FP4	0.916 ± 0.003	0.582 ± 0.007	Ø	0.573 ± 0.013	0.954 ± 0.003	0.943 ± 0.005
	Graph	0.920 ± 0.003	0.582 ± 0.007	Ø	0.573 ± 0.013	0.940 ± 0.003	0.948 ± 0.005
	KR	0.932 ± 0.004	0.582 ± 0.007	Ø	0.573 ± 0.013	0.971 ± 0.002	0.953 ± 0.005
	MACCS	0.919 ± 0.004	0.582 ± 0.007	Ø	0.573 ± 0.013	0.941 ± 0.003	0.950 ± 0.005
	PubChem	0.916 ± 0.003	0.582 ± 0.007	Ø	0.573 ± 0.013	0.937 ± 0.003	0.949 ± 0.005
NRs	CDK	0.860 ± 0.003	0.754 ± 0.013	Ø	0.775 ± 0.014	0.872 ± 0.006	0.916 ± 0.005
	CDKExt	0.859 ± 0.004	0.754 ± 0.013	Ø	0.775 ± 0.014	0.871 ± 0.006	0.916 ± 0.005
	FP4	0.859 ± 0.004	0.754 ± 0.013	Ø	0.775 ± 0.014	0.906 ± 0.005	0.911 ± 0.005
	Graph	0.857 ± 0.005	0.754 ± 0.013	Ø	0.775 ± 0.014	0.878 ± 0.006	0.905 ± 0.006
	KR	0.879 ± 0.008	0.754 ± 0.013	Ø	0.775 ± 0.014	0.932 ± 0.005	0.915 ± 0.005
	MACCS	0.857 ± 0.004	0.754 ± 0.013	Ø	0.775 ± 0.014	0.881 ± 0.006	0.912 ± 0.005
	PubChem	0.855 ± 0.004	0.754 ± 0.013	Ø	0.775 ± 0.014	0.876 ± 0.006	0.912 ± 0.005
Global	CDK	0.849 ± 0.005	0.916 ± 0.006	Ø	0.924 ± 0.005	0.909 ± 0.005	0.936 ± 0.004
	CDKExt	0.849 ± 0.005	0.916 ± 0.006	Ø	0.924 ± 0.005	0.907 ± 0.005	0.936 ± 0.004
	FP4	0.843 ± 0.004	0.916 ± 0.006	Ø	0.924 ± 0.005	0.949 ± 0.004	0.935 ± 0.004
	Graph	0.850 ± 0.005	0.916 ± 0.006	Ø	0.924 ± 0.005	0.921 ± 0.005	0.935 ± 0.004
	KR	0.862 ± 0.002	0.916 ± 0.006	Ø	0.924 ± 0.005	0.947 ± 0.004	0.936 ± 0.004
	MACCS	0.851 ± 0.007	0.916 ± 0.006	Ø	0.924 ± 0.005	0.937 ± 0.004	0.936 ± 0.004
	PubChem	0.848 ± 0.005	0.916 ± 0.006	Ø	0.924 ± 0.005	0.923 ± 0.005	0.936 ± 0.004

Ø represents the fact that we did not compute the prediction performance because of data reason; ∗ stands for the prediction results derived from previous studies.

4. Experiments and Results

4.1. Benchmark Evaluation and Evaluation Indices. In order to demonstrate the performance of our method, we adopt the 10-fold cross validation and external validation. The 10-fold validation was widely used in prediction of DTIs [29, 48, 49] and other interaction prediction in bioinformatics. The main experiment process is that the whole dataset is randomly divided into 10 groups; each group alternates as a testing set, and the rest of the 9 groups alternate as the training set, and this process is repeated 10 times.

Furthermore, the DTIs of new chemical entities and failed drugs are a very important portion in this study. We use two external datasets (ExGPCRs, ExKinases) to evaluate performance of our method by predicting all interactions with them.

We use the AUC (area under the ROC curve) as an evaluation metric for our SDTRLS as for SDTNBI methods, and the values in Tables 3, 5, and 6 are presented in the format

Table 4: The performance of two external validations.

		AUC					
Target	FP	DBSI-R	NWNBI	EWNBI*	NBI	SDTNBI	SDTRLS
ExGPCRs	CDK	0.752	0.756	0.764	0.769	0.753	0.824
	CDKExt	0.751	0.756	0.764	0.769	0.751	0.804
	FP4	0.758	0.756	0.764	0.769	0.784	0.818
	Graph	0.758	0.756	0.764	0.769	0.761	0.842
	KR	0.770	0.756	0.764	0.769	0.797	0.840
	MACCS	0.754	0.756	0.764	0.769	0.758	0.822
	PubChem	0.754	0.756	0.764	0.769	0.759	0.831
ExKinases	CDK	0.851	0.812	0.821	0.828	0.852	0.827
	CDKExt	0.850	0.812	0.821	0.828	0.852	0.834
	FP4	0.850	0.812	0.821	0.828	0.847	0.855
	Graph	0.851	0.812	0.821	0.828	0.852	0.841
	KR	0.851	0.812	0.821	0.828	0.863	0.848
	MACCS	0.851	0.812	0.821	0.828	0.852	0.844
	PubChem	0.850	0.812	0.821	0.828	0.852	0.846

∗ stands for the prediction results derived from previous studies.

of mean ± standard deviation. The larger the AUC value is, the better the prediction is.

4.2. Cross Validation. Table 3 describes the performance evaluation index values of the predicted datasets in the 10-fold cross validation for 5 datasets. SDTRLS's minimum AUC among the seven substructures reaches 0.979 and the average is 0.981, which indicates good prediction results. However, on NRs dataset, the validation results of each substructure are relatively poor and the minimum value is 0.905 based on Graph substructure while the maximum value is 0.916. On Kinases dataset, the verification results are also very stable, with the maximum and minimum values of 0.973 and 0.969, respectively. On ICs dataset, the verification results are not bad, the minimum value of AUC is 0.943 with FP4 substructure, and the maximum value is 0.956 with CDK substructure. Similarly, on Global dataset, the results are stable except for the slightly lower values of 7 substructures, between 0.935 and 0.936. In general, the validation results on GPCR and Kinase datasets are better than the other three datasets. Moreover, the prediction performances of EWNBI, NWNBI, and NBI on Kinases dataset are lightly better than SDTRLS, while SDTRLS has obvious advantage on ICs, NRs, and Global datasets. In addition, because the authors did not provide the data needed for EWNBI method on three datasets (ICs, NRs, and Global) and the prediction results of datasets GPCRs and Kinases are not good, we do not compute the AUC values of EWNBI method on these three datasets. Overall, SDTRLS and SDTNBI provide more stable prediction results on 5 datasets.

4.3. External Validation. Table 4 describes the evaluation results of six methods on two external datasets ExGPCRs and ExKinases; the basic datasets are GPCRs and Kinases, respectively. Overall, external validation results of all prediction methods are worse than 10-fold cross validation results because new chemical entities have no known DTIs. On ExGPCRs dataset, the AUC values of SDTRLS on the 7 substructures are between 0.804 and 0.842. On ExKinases dataset, the AUC values of SDTRLS of the 7 substructures are between 0.827 and 0.855. As can be seen from Table 4, the verification results of all approaches on ExKinases are better than on ExGPCRs. In the validation on ExKinases dataset, there are no obvious differences in AUC values among DBSI-R, SDTNBI, and SDTRLS. On ExKinases, SDTRLS demonstrates its excellent prediction power.

4.4. Comparison with Previous Methods. Since the datasets used in this study are derived from the datasets used in the SDTNBI method, as the state-of-the-art method, its prediction performances are more stable than the other 4 methods. In this study, the comparison is performed in terms of the t-test statistical analyses of SDTRLS and SDTNBI methods, as well as in terms of the parameter-independent AUC value with other 5 methods.

Table 5 shows t-tests results of SDTNBI and SDTRLS on five datasets GPCRs, Kinases, ICs, NRs, and Global, respectively. We can see from Table 5 that the average AUC of our method on each dataset is greater than that of the SDTNBI method, especially in the GPCRs and Kinases datasets, respectively, from 0.928 to 0.981 and from 0.919 to 0.971. Moreover, there were significant differences ($p < 0.05$) in the comparison results of GPCRs, Kinases, and NRs datasets; particularly, the comparison result is more significant ($p < 0.01$) on GPCRs and Kinases datasets. In conclusion, our method is more stable than the SDTNBI method in terms of the 10-fold cross validation.

We also compare the prediction results with other four methods on five datasets GPCRs, Kinases, ICs, NRs, and Global. The four competing methods are NBI, NWNBI, EWNBI, and DBSI-R [29]. NBI applied a mass diffusion-based method to obtain the predicted list by considering

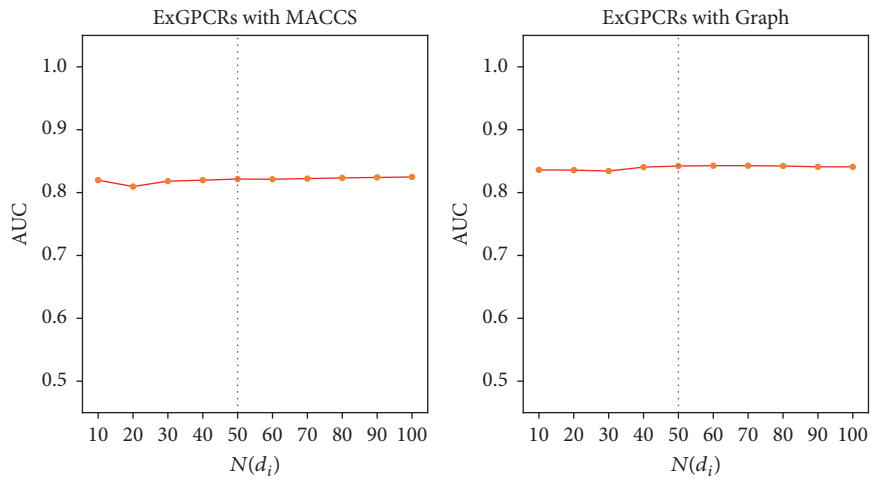

Figure 1: Robustness of SDTRLS with respect to the number of $N(d_i)$: the dotted line is the default value and its prediction performance.

Table 5: The t-tests results of 10-fold cross validations on 5 datasets.

	AUC				
Methods	GPCRs	Kinases	ICs	NRs	Global
SDTNBI	0.928 ± 0.026	0.919 ± 0.028	0.944 ± 0.014	0.888 ± 0.023	0.928 ± 0.017
SDTRLS	0.981 ± 0.001	0.971 ± 0.001	0.951 ± 0.005	0.912 ± 0.004	0.935 ± 0.000
p	0.0002	0.0004	0.248	0.016	0.232

Table 6: The t-tests results of two external validations.

	AUC	
Methods	ExGPCRs	ExKinases
SDTNBI	0.766 ± 0.017	0.853 ± 0.005
SDTRLS	0.825 ± 0.013	0.842 ± 0.009
p	$1.04e-05$	0.019

the bipartite graph. However, in EWNBI method, a DTI network was weighted by the potency of binding affinity or inhibitory activity of the interactions with drugs and targets. The theoretical basis of NWNBI method is that the hub node is more difficult to be influenced. The DBSI method is based on the hypothesis that two similar drugs may have similar targets. Table 3 shows that the SDTRLS method is slightly better than NBI, NWNBI, and EWNBI methods on GPCRs and much better than DBSI-R method. In addition, SDTRLS method is much better than DBSI-R method while being comparable with NBI, NWNBI, and EWNBI methods on Kinases. In general, the SDTRLS approach is comparable to these four methods from the results of the 10-fold cross validation on GPCRs and Kinases datasets.

Table 6 shows results of SDTNBI and SDTRLS on two datasets, ExGPCRs and ExKinases, respectively. From Table 6 we can see that our method greatly outperforms the SDTNBI method, on ExGPCRs in terms of the average AUC and t-test result ($p < 0.01$). In addition, the average AUC of our methods are slightly lower than those of SDTNBI method on ExKinases, which may be due to the sparsity of known DTIs in this dataset.

We compare the prediction result of our method with other four competing methods on the same datasets ExGPCRs and ExKinases. We can see from Table 4 that SDTRLS method outperforms the other four competing methods on ExGPCRs dataset. In addition, SDTRLS method is also comparable with other four competing methods on ExKinases dataset.

4.5. Parameter Analysis for $N(d_i)$ and K. In this section, we analyzed two parameters, including $N(d_i)$ for similarity network fusion and K for new chemical entities. The parameter h was set to be 1 according to previous study [27]. Moreover, GIP is widely used in other studies [10, 34, 37, 50, 51]; we also set the values of both α and β to be 1. All results were validated over external validation of ExGPCRs datasets based on substructures MACCS and Graph. Figure 1 describes that the sensitivity of the prediction performance of SDTRLS with to different numbers $N(d_i)$ of similarity network fusion. SDTRLS had stable prediction performance over a wide range from 10 to 100. The impact of parameter K for new chemical entities on the prediction performance of SDTRLS, in terms of AUC value, is illustrated in Figure 2. SDTRLS was robust to different values of parameter K.

4.6. Case Studies. In order to further confirm the prediction ability of our method, we conduct an experimental analysis

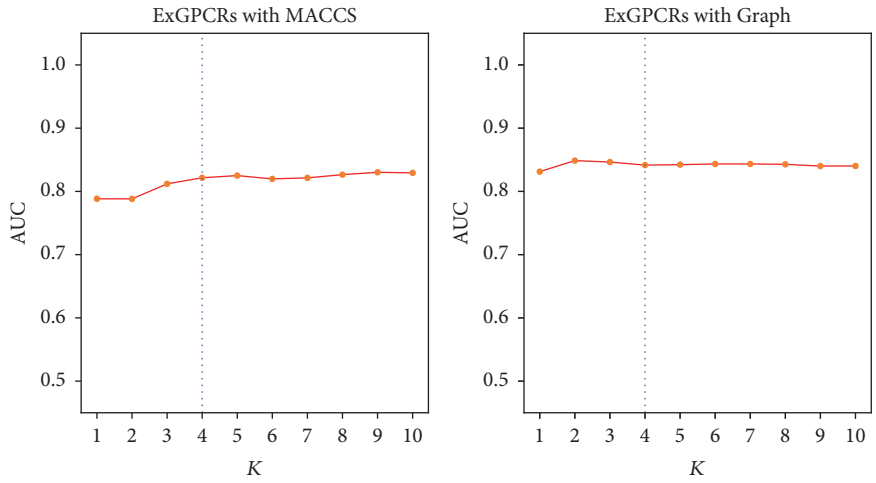

Figure 2: Robustness of SDTRLS with respect to the number of K: the dotted line is the default value and its prediction performance.

on dataset ExGPCRs, and its known DTIs are not used as a priori knowledge when conducting external validation. The selected predictions of drugs are confirmed with DrugBank, ChEMBL, and KEGG databases. Table 7 describes the confirmed result based on ExGPCRs dataset. We select the top five predicted interactions of 5 drugs; the top one predicted interaction of every drug is confirmed by searching databases. Furthermore, 76% of all predicted DTIs (19 out of 25) are also confirmed with three databases; 32% of predicted DTIs (8 out of 25) are simultaneously confirmed with two databases, especially in the predicted result of (DB00209, DB00283, and DB00334); they are all confirmed with the several databases. In addition, we further validate the results marked as unknown in the prediction results; we searched the relevant literature and found the related description. For example, thiethylperazine (DB00372) is an antagonist of human Dopamine D3 (hDRD3_170) according to the description in Petsko and Ringe [52] which shows that the prediction result is meaningful, and other remaining unknown DTIs deserve being validated in the future. In general, it proves that our method is effective in practical applications.

5. Conclusions

The systematic understanding of the interactions between chemical compounds and target proteins is very important for new drug design and development. In the past decades, in order to solve the time-consuming shortcomings of traditional biochemical methods, many computational approaches have been developed to predict DTIs, like machine learning, network inference, and so on. However, these methods mainly focused on new DTIs for known drugs and paid less attention to new chemical entities for DTIs. In addition, their prediction performances are not good enough.

In this study, we have constructed the similarity kernel of approved drugs, failed drugs, and new chemical entities

Table 7: The new confirmation of drug-target interactions based on Graph substructure in the ExGPCRs.

Drug ID	Target ID	Rank	Source
DB00209	hCHRM2_86	1	KEGG
	hCHRM3_98	2	KEGG
	hCHRM1_92	3	DrugBank KEGG
	hCHRM4_87	4	KEGG
	hCHRM5_90	5	KEGG
DB00283	hDRD3_170	1	ChEMBL
	hDRD4_106	2	ChEMBL
	HDRD2_94	3	ChEMBL
	hOPRM1_166	4	ChEMBL
	hOPRK1_173	5	ChEMBL
DB00334	hCHRM2_86	1	DrugBank ChEMBL
	hCHRM3_98	2	DrugBank ChEMBL
	hCHRM1_92	3	DrugBank ChEMBL
	hA1AB_164	4	ChEMBL KEGG
	hA1AD_116	5	ChEMBL KEGG
DB00372	hDRD2_94	1	DrugBank KEGG
	h5HT2A_125	2	Unknown
	h5HT2C_126	3	Unknown
	hDRD3_170	4	Unknown
	h5HT1A_89	5	Unknown
DB00612	hB1AR_88	1	DrugBank KEGG
	hB2AR_84	2	DrugBank
	hB3AR_93	3	Unknown
	hDRD2_94	4	Unknown
	hOPRM1_166	5	Unknown

by weighting the chemical substructures. Then, GIP kernels were calculated from drugs and targets according to the known DTIs. For the new chemical entities or failed drugs, we used the KNN to initialize the DTIs before calculating the GIP kernel. To construct a comprehensive similarity kernel for drugs, SNF method is used to fuse GIP kernel and substructure similarity kernel. Finally, the score of drug-target pairs was predicted by Kron_RLS. We compared the prediction performance with other competing methods via the tenfold cross validation and external validation.

However, there are still some limitations in this study. First, since the target set is specified within the current datasets, it may be not possible to predict the DTIs of the target beyond the datasets. Other similarity information of targets such as the sequence and functional network [53–56] is not used when the similarity kernel of targets is constructed. In addition, the 3D structure of drugs may also need to be considered as important information. It is expected that additional information may improve prediction performance. In the future, more information using other methods such as ClusterViz [57] should be integrated to develop a more efficient prediction method. Nevertheless, this study provides an important basis for new drug development and drug repositioning and also plays an important role in the personalized medical development.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work has been supported in part by the National Natural Science Foundation of China under Grants no. 61772552, no. 61420106009, and no. 61622213.

References

[1] T. T. Ashburn and K. B. Thor, "Drug repositioning: identifying and developing new uses for existing drugs," *Nature Reviews Drug Discovery*, vol. 3, no. 8, pp. 673–683, 2004.

[2] N. Novac, "Challenges and opportunities of drug repositioning," *Trends in Pharmacological Sciences*, vol. 34, no. 5, pp. 267–272, 2013.

[3] M. R. Hurle, L. Yang, Q. Xie, D. K. Rajpal, P. Sanseau, and P. Agarwal, "Computational drug repositioning: from data to therapeutics," *Clinical Pharmacology & Therapeutics*, vol. 93, no. 4, pp. 335–341, 2013.

[4] A. L. Hopkins, "Drug discovery: predicting promiscuity," *Nature*, vol. 462, no. 7270, pp. 167-168, 2009.

[5] S. J. Swamidass, "Mining small-molecule screens to repurpose drugs," *Briefings in Bioinformatics*, vol. 12, no. 4, pp. 327–335, 2011.

[6] B. Y. Feng, A. Simeonov, A. Jadhav et al., "A high-throughput screen for aggregation-based inhibition in a large compound library," *Journal of Medicinal Chemistry*, vol. 50, no. 10, pp. 2385–2390, 2007.

[7] M. A. Yildirim, K.-I. Goh, M. E. Cusick, A.-L. Barabási, and M. Vidal, "Drug-target network," *Nature Biotechnology*, vol. 25, no. 10, pp. 1119–1126, 2007.

[8] S. Zhao and S. Li, "Network-based relating pharmacological and genomic spaces for drug target identification," *PLoS ONE*, vol. 5, no. 7, Article ID e11764, 2010.

[9] S. Alaimo, A. Pulvirenti, R. Giugno, and A. Ferro, "Drug-target interaction prediction through domain-tuned network-based inference," *Bioinformatics*, vol. 29, no. 16, pp. 2004–2008, 2013.

[10] J.-P. Mei, C.-K. Kwoh, P. Yang, X.-L. Li, and J. Zheng, "Drug-target interaction prediction by learning from local information and neighbors," *Bioinformatics*, vol. 29, no. 2, pp. 238–245, 2013.

[11] T. van Laarhoven and E. Marchiori, "Predicting Drug-Target Interactions for New Drug Compounds Using a Weighted Nearest Neighbor Profile," *PLoS ONE*, vol. 8, no. 6, Article ID e66952, 2013.

[12] Q. Yuan, J. Gao, D. Wu, S. Zhang, H. Mamitsuka, and S. Zhu, "DrugE-Rank: improving drug-target interaction prediction of new candidate drugs or targets by ensemble learning to rank," *Bioinformatics*, vol. 32, no. 12, pp. i18–i27, 2016.

[13] H. Luo, J. Wang, M. Li et al., "Drug repositioning based on comprehensive similarity measures and Bi-Random walk algorithm," *Bioinformatics*, vol. 32, no. 17, pp. 2664–2671, 2016.

[14] E. E. Bolton, Y. Wang, P. A. Thiessen, and S. H. Bryant, Chapter 12 - PubChem: Integrated Platform of Small Molecules and Biological Activities. Elsevier Science & Technology, 2008.

[15] D. S. Wishart, C. Knox, A. C. Guo et al., "DrugBank: a knowledgebase for drugs, drug actions and drug targets," *Nucleic Acids Research*, vol. 36, pp. D901–D906, 2008.

[16] A. Gaulton, L. J. Bellis, A. P. Bento et al., "ChEMBL: a large-scale bioactivity database for drug discovery," *Nucleic Acids Research*, vol. 40, no. 1, pp. D1100–D1107, 2012.

[17] C. Qin, C. Zhang, F. Zhu et al., "Therapeutic target database update 2014: a resource for targeted therapeutics," *Nucleic Acids Research*, vol. 42, no. 1, pp. D1118–D1123, 2014.

[18] M. Kanehisa and S. Goto, "KEGG: kyoto encyclopedia of genes and genomes," *Nucleic Acids Research*, vol. 28, no. 1, pp. 27–30, 2000.

[19] M. Kuhn, M. Campillos, I. Letunic, L. J. Jensen, and P. Bork, "A side effect resource to capture phenotypic effects of drugs," *Molecular Systems Biology*, vol. 6, p. 343, 2010.

[20] M. Kuhn, D. Szklarczyk, S. Pletscher-Frankild et al., "STITCH 4: integration of protein-chemical interactions with user data," *Nucleic Acids Research*, vol. 42, no. 1, pp. D401–D407, 2014.

[21] A. Franceschini, D. Szklarczyk, S. Frankild et al., "STRING v9.1: protein-protein interaction networks, with increased coverage and integration," *Nucleic Acids Research*, vol. 41, no. 1, pp. D808–D815, 2013.

[22] T. Liu, Y. Lin, X. Wen, R. N. Jorissen, and M. K. Gilson, "BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities," *Nucleic Acids Research*, vol. 35, supplement 1, pp. D198–D201, 2007.

[23] J. T. Dudley, T. Deshpande, and A. J. Butte, "Exploiting drug-disease relationships for computational drug repositioning," *Briefings in Bioinformatics*, vol. 12, no. 4, pp. 303–311, 2011.

[24] H. Ding, I. Takigawa, H. Mamitsuka, and S. Zhu, "Similarity-basedmachine learning methods for predicting drug-target interactions: A brief review," *Briefings in Bioinformatics*, vol. 15, no. 5, pp. 734–747, 2013.

[25] Y. Tabei and Y. Yamanishi, "Scalable prediction of compound-protein interactions using minwise hashing.," *BMC systems biology*, vol. 7, p. S3, 2013.

[26] H. Yabuuchi, S. Niijima, H. Takematsu et al., "Analysis of multiple compound-protein interactions reveals novel bioactive molecules," *Molecular Systems Biology*, vol. 7, article no. 472, 2011.

[27] Y. Yamanishi, M. Kotera, M. Kanehisa, and S. Goto, "Drug-target interaction prediction from chemical, genomic and pharmacological data in an integrated framework," *Bioinformatics*, vol. 26, no. 12, Article ID btq176, pp. i246–i254, 2010.

[28] X. Chen, C. C. Yan, X. Zhang et al., "Drug-target interaction prediction: Databases, web servers and computational models," *Briefings in Bioinformatics*, vol. 17, no. 4, pp. 696–712, 2016.

[29] F. Cheng, C. Liu, J. Jiang et al., "Prediction of drug-target interactions and drug repositioning via network-based inference," *PLoS Computational Biology*, vol. 8, no. 5, Article ID e1002503, 2012.

[30] X. Chen, M.-X. Liu, and G.-Y. Yan, "Drug-target interaction prediction by random walk on the heterogeneous network," *Molecular BioSystems*, vol. 8, no. 7, pp. 1970–1978, 2012.

[31] K. Bleakley, G. Biau, and J.-P. Vert, "Supervised reconstruction of biological networks with local models," *Bioinformatics*, vol. 23, no. 13, pp. i57–i65, 2007.

[32] F. Mordelet and J.-P. Vert, "SIRENE: Supervised inference of regulatory networks," *Bioinformatics*, vol. 24, no. 16, pp. i76–i82, 2008.

[33] K. Bleakley and Y. Yamanishi, "Supervised prediction of drug-target interactions using bipartite local models," *Bioinformatics*, vol. 25, no. 18, pp. 2397–2403, 2009.

[34] T. van Laarhoven, S. B. Nabuurs, and E. Marchiori, "Gaussian interaction profile kernels for predicting drug-target interaction," *Bioinformatics*, vol. 27, no. 21, pp. 3036–3043, 2011.

[35] W. Lan, J. Wang, M. Li et al., "Predicting drug–target interaction using positive-unlabeled learning," *Neurocomputing*, vol. 206, pp. 50–57, 2016.

[36] W. Lan, J. Wang, M. Li, J. Liu, F. X. Wu, and Y. Pan, "Predicting microrna-disease associations based on improved microrna and disease similarities," *IEEEACM Transactions on Computational Biology & Bioinformatics*, vol. PP, no. 99, p. 1, 2016.

[37] C. Yan, J. Wang, P. Ni, W. Lan, F. X. Wu, and Y. Pan, "Dnrlmf-mda:predicting microrna-disease associations based on similarities of micrornas and diseases," *IEEEACM Transactions on Computational Biology & Bioinformatics*, 2017.

[38] M. Gönen, "Predicting drug-target interactions from chemical and genomic kernels using Bayesian matrix factorization," *Bioinformatics*, vol. 28, no. 18, pp. 2304–2310, 2012.

[39] X. Zheng, H. Ding, H. Mamitsuka, and S. Zhu, "Collaborative matrix factorization with multiple similarities for predicting drug-target interactions," in *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1025–1033, Chicago, Ill, USA, August 2013.

[40] A. Ezzat, P. Zhao, M. Wu, and X. Li, "Drug-target interaction prediction with graph regularized matrix factorization," *IEEEACM Transactions on Computational Biology & Bioinformatics*, vol. 1, p. 1, 2016.

[41] M. Hay, D. W. Thomas, J. L. Craighead, C. Economides, and J. Rosenthal, "Clinical development success rates for investigational drugs," *Nature Biotechnology*, vol. 32, no. 1, pp. 40–51, 2014.

[42] A. Mullard, "Drug repurposing programmes get lift off," *Nature Reviews Drug Discovery*, vol. 11, no. 7, pp. 505-506, 2012.

[43] Z. Wu, F. Cheng, J. Li, W. Li, G. Liu, and Y. Tang, "SDTNBI: an integrated network and chemoinformatics tool for systematic prediction of drug–target interactions and drug repositioning," *Briefings in Bioinformatics*, pp. 333–347, 2016.

[44] B. Wang, A. M. Mezlini, F. Demir et al., "Similarity network fusion for aggregating data types on a genomic scale," *Nature Methods*, vol. 11, no. 3, pp. 333–337, 2014.

[45] J. Basilico and T. Hofmann, "Unifying collaborative and content-based filtering," in *Proceedings of the Proceedings, Twenty-First International Conference on Machine Learning, ICML 2004*, pp. 65–72, July 2004.

[46] A. Ben-Hur and W. S. Noble, "Kernel methods for predicting protein-protein interactions," *Bioinformatics*, vol. 21, no. 1, pp. i38–i46, 2005.

[47] M. Hue and J.-P. Vert, "On learning with kernels for unordered pairs," in *Proceedings of the 27th International Conference on Machine Learning, ICML 2010*, pp. 463–470, June 2010.

[48] Z. Xia, L. Wu, X. Zhou, and S. T. Wong, "Semi-supervised drug-protein interaction prediction from heterogeneous biological spaces," *BMC Systems Biology*, vol. 4, no. Suppl 2, p. S6, 2010.

[49] C. Huang, R. Zhang, Z. Chen et al., "Predict potential drug targets from the ion channel proteins based on SVM," *Journal of Theoretical Biology*, vol. 262, no. 4, pp. 750–756, 2010.

[50] Z.-H. You, Z.-A. Huang, Z. Zhu et al., "Pbmda: A novel and effective path-based computational model for mirna-disease association prediction," *PLoS Computational Biology*, vol. 13, no. 3, 2017.

[51] W. Lan, M. Li, K. Zhao et al., "LDAP: a web server for lncRNA-disease association prediction," *Bioinformatics*, vol. 33, no. 3, pp. 458–460, 2016.

[52] G. Petsko and D. Ringe, "ligand-based virtual screening for new antagonists of dopamine receptor D2/D3," *Shanghai Management Science*, 2013.

[53] X. Peng, J. Wang, W. Peng, F. Wu, and Y. Pan, "Protein–protein interactions: detection, reliability assessment and applications," *Briefings in Bioinformatics*, vol. 18, no. 5, pp. 798–819, 2016.

[54] B. Zhao, J. Wang, and F. Wu, "Computational Methods to Predict Protein Functions from Protein-Protein Interaction Networks," *Current Protein & Peptide Science*, vol. 18, no. 11, 2017.

[55] B. Zhao, J. Wang, X. Li, and F.-X. Wu, "Essential protein discovery based on a combination of modularity and conservatism," *Methods*, vol. 110, pp. 54–63, 2016.

[56] B. Zhao, J. Wang, M. Li et al., "A New Method for Predicting Protein Functions from Dynamic Weighted Interactome Networks," *IEEE Transactions on NanoBioscience*, vol. 15, no. 2, pp. 133–141, 2016.

[57] J. Wang, J. Zhong, G. Chen, M. Li, F.-X. Wu, and Y. Pan, "ClusterViz: A Cytoscape APP for Cluster Analysis of Biological Network," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 12, no. 4, pp. 815–822, 2015.