BMC Bioinformatics

DDIGIP: predicting drug-drug interactions based on Gaussian interaction profile kernels

Cheng Yan[1,2], Guihua Duan[1], Yi Pan[3], Fang-Xiang Wu[4] and Jianxin Wang[1*]

Abstract

Background: A drug-drug interaction (DDI) is defined as a drug effect modified by another drug, which is very common in treating complex diseases such as cancer. Many studies have evidenced that some DDIs could be an increase or a decrease of the drug effect. However, the adverse DDIs maybe result in severe morbidity and even morality of patients, which also cause some drugs to withdraw from the market. As the multi-drug treatment becomes more and more common, identifying the potential DDIs has become the key issue in drug development and disease treatment. However, traditional biological experimental methods, including in vitro and vivo, are very time-consuming and expensive to validate new DDIs. With the development of high-throughput sequencing technology, many pharmaceutical studies and various bioinformatics data provide unprecedented opportunities to study DDIs.

Result: In this study, we propose a method to predict new DDIs, namely DDIGIP, which is based on Gaussian Interaction Profile (GIP) kernel on the drug-drug interaction profiles and the Regularized Least Squares (RLS) classifier. In addition, we also use the k-nearest neighbors (KNN) to calculate the initial relational score in the presence of new drugs via the chemical, biological, phenotypic data of drugs. We compare the prediction performance of DDIGIP with other competing methods via the 5-fold cross validation, 10-cross validation and de novo drug validation.

Conlusion: In 5-fold cross validation and 10-cross validation, DDRGIP method achieves the area under the ROC curve (AUC) of 0.9600 and 0.9636 which are better than state-of-the-art method (L1 Classifier ensemble method) of 0.9570 and 0.9599. Furthermore, for new drugs, the AUC value of DDIGIP in de novo drug validation reaches 0.9262 which also outperforms the other state-of-the-art method (Weighted average ensemble method) of 0.9073. Case studies and these results demonstrate that DDRGIP is an effective method to predict DDIs while being beneficial to drug development and disease treatment.

Keywords: Drug, Drug-drug interaction, Gaussian interaction profile, RLS

*Correspondence: jxwang@mail.csu.edu.cn
[1]School of Computer Science and Engineering, Central South University, 932
South Lushan Rd, 410083, ChangSha, China
Full list of author information is available at the end of the article

Yan *et al. BMC Bioinformatics* 2019, **20**(Suppl 15):538

Page 2 of 10

Background

Drug-drug interactions (DDI) is defined as that a drug affects the efficacy of another drug when multi-drugs are adopted in the treatment of a disease [1]. DDIs can lead to change systemic exposure and result in variations in drug responses, which can improve the drugs efficiency and the life quality of illnesses [2]. However, DDIs also can cause serious adverse effects, drug withdrawal from the market and even the patient morality [3, 4]. Meanwhile, with the medical technology development and personalized medical requirements, more and more patients were simultaneously treated by multi-drugs and between 2009 and 2012, 38.1% of U.S. adults aging 18-44 years used three or more prescription drugs during a 30 day time period [5–7]. Therefore, identifying the potential DDIs has become a major issue in drug development and practice process.

With the high-throughput sequencing technology development, many databases related to drugs have been constructed. DrugBank database can provide drug targets, drug enzymes, drug transporters and DDIs, which are widely used in studying drug-target associations and drug repositioning [8–10]. PubChem Compound database contains the chemical substructures and their biological test results [11]. In addition, SIDER and OFFSIDES databases include drug side effects and "off-label" side effects, respectively [12, 13]. KEGG database contains drug pathways and chemical substructures [14]. TWOSIDES database contains the DDIs based on the adverse event reports in the AERS (adverse effect reactions) [13, 15].

The above mentioned databases related to drugs were extracted from the published literature and reports with experimental validation, and could provide the basis to the development of computational methods to predict new DDIs. Recently, many computational methods have been proposed to predict potential DDIs based on the assumption that similar drugs tend to interact with similar other drugs. These approaches usually used the biological network data, chemical substructure data and phenotypic data. Based on MACCS substructures of drugs, Vilar et al. developed a similarity-based model to predict new DDIs [16]. Liu et al. proposed a model to predict potential DDIs via random forest-based classification model, which also adopted a feature selection technique over the chemical substructures, protein-protein interactions between targets of drugs and target enrichment of KEGG pathways [17]. Cheng et al. proposed a method to infer novel DDIs via machine learning classifiers, whose major feature is integrating drug chemical, phenotypic and genomic properties [18]. IPFs (interaction profile fingerprints) method was proposed to predict hidden DDIs [19]. Logistic regression model was used to predict new DDIs by Takeda et al., which analyzed

the effects of 2D structural similarities of drugs on DDI prediction with other pharmacokinetics (PK) and pharmacodynamics (PD) knowledge [20]. Via constructing the drug similarity based on their 2D and 3D molecular substructures, targets, side effects and known DDIs, Vilar et al. further proposed a method to predict new DDIs on a large scale data, where the key feature is capturing the characteristics of drugs by 3D substructures when 2D substructures are missing [21]. Herrero-Zazo et al. provided a computational method to predict DDIs by different types of DDIs and their mechanisms [22]. By integrating similarities from drug molecular and pharmacological phenotypes, Li et al. used a Bayesian network to provide large-scale exploration and analysis of drug combinations [23]. Through calculating the functional similarity from drug carriers, drug transporters, drug enzymes and drug targets, Ferdousi et al. developed an approach to discover new DDIs [24]. Based on the Probabilistic Soft Logic method, a computational framework was developed to discover new DDIs by integrating the multiple drug similarities and known DDIs [25]. The label propagation approach was also developed to discover new DDIs, which used drug chemical structures, side effects and off side effects [26]. In order to predict drug adverse drug reactions (ADRs), a systems pharmacology model called MEF (multiple evidence fusion) has been developed by integrating known DDIs and other similarities of drugs [27]. Based on the assumption that synergistic effects of drugs are usually similar, Network-based Laplacian regularized Least Square Synergistic (NLLSS) method was developed to predict novel DDIs [28]. Via calculating the similarities of chemical, biological, phenotypic and known DDIs of drugs, Zhang et al. proposed three ensemble methods to predict novel DDIs, which included a weight average ensemble method and two classifier ensemble methods (L1 and L2) [29].

In addition, many other approaches used quantitative structure-activity relationship (QSAR) model, clinical data and data mining to study DDIs. Matthews et al. developed 14 QSAR models to predict the cardiac adverse effects for generic pharmaceutical substances [30]. Zakharov et al. developed QSAR models to predict the likelihood of DDIs for any pair of drugs by radial basis functions with self-consistent regression (RBF-SCR) and random forest (RF) [31]. Cami et al. proposed a Predictive Pharmacointeraction Networks (PPINs) to predict novel DDIs by exploiting the known DDIs and other intrinsic and taxonomic properties of drugs and AEs [32]. Huang et al. developed a method to predict DDIs using protein-protein interaction network and clinical side effects [33]. Based on information of drug metabolism, text-mining and reasoning methods were developed to infer new DDIs [34]. Iyer et al. used the textual portion Electronic health records (EHRs) to directly discover new DDIs [35]. Banda

Yan *et al. BMC Bioinformatics* 2019, **20**(Suppl 15):538

Page 3 of 10

et al. also adopted a data mining method to predict new DDIs from the EHRs [36]. Based on the k-nearest neighbor algorithm, Chen et al. proposed a model to predict DDIs which integrated nine predictors by majority voting [37]. Furthermore, the drug response prediction and drug-target interaction prediction are also the important research topics about drugs. By integrating genomic/pharmaceutical data, protein-protein interaction network, and prior knowledge of drug-target interactions with the techniques of network propagation, Wang et al. have developed a dNetFS method to prioritize genetic and gene expression features of cancer cell lines that predict drug response [38]. Based on the massively collected drug-kinase interactions and drug sensitivity datasets, Liu et al. employed a sparse linear model to infer essential kinases governing the cellular responses to drug treatments in cancer cells [39]. Based on the sequence information of both targets and drugs, DeepDTA is used to predict drug-target interaction binding affinities, which is a deep-learning based model (convolutional neural networks) [40].

Although the above DDI prediction methods have achieved some good prediction results of novel DDIs and provided useful information for drug development and practice process. However, these methods did not pay enough attention to new drugs which do not have any DDIs with other drugs or cannot predict novel DDIs for new drugs because known DDIs are missing.

In this study, we develop a computational method (called DDIGIP) to predict novel DDIs based on drug Gaussian interaction profile (GIP) kernel similarity and regularized least squares (RLS) classifier. We calculate the GIP similarity of drugs by known DDIs, and then adopt the RLS method to compute the related scores of any drug pairs. In addition, when predicting DDIs for new drugs, we use the KNN method to compute the initial relational scores by similarity calculated from some important chemical, biological and phenotypic information of drugs. The drug chemical structures, drug-target interactions, drug enzymes, drug transports, drug pathways, drug indications, drug side effects and drug off side effects are all used to calculate similarity of drugs. 5-fold cross validation (5CV), 10-fold cross validation (10CV) and de novo drug validation are used to systemically assess prediction performance of DDIGIP, compared with other methods. In 5-fold cross validation, the area under the ROC curve (AUC) value of DDIGIP is 0.9600 which is slightly better than the state-of-the-art method L1 classifier ensemble (L1E) method results of 0.9570. In addition, the experimental results of 10-fold cross validation also demonstrate that DDIGIP outperforms the L1E method. In de novo drug validation, DDIGIP achieves the AUC of 0.9262, which is also better than the weighted average ensemble (WAE) method result of 0.9073. Case

studies further validate the prediction ability of DDIGIP method.

Materials

In this study, the benchmark dataset of DDIs composes of 548 drugs and 48,584 DDIs. This dataset is obtained from the TWOSIDES database. In addition, because we need to calculate the relational scores of new drugs, we also download other chemical, biological and phenotypic data from other databases to compute the similarity of drugs. The chemical data are PubChem substructures which are downloaded from the PubChem Compound database. Biological data include drug targets, drug transports, drug enzymes and drug pathways, the first three types are obtained from the DrugBank database and the last one is from the KEGG database. Furthermore, the phenotypic data composes of drug indications, drug side effects and drug off side effects. The SIDER database provided the drug indications and drug side effects, and OFFSIDES provided the drug off side effects.

Previous studies also provided the download links for these datasets [29]. Table 1 shows the relevant information which includes data type, data source and dimensionality.

Methods

GIP kernel similarity of drugs

The GIP kernel similarity has widely been used in other prediction issues of similar areas and achieved effective prediction performances [41–46]. RLS-Kron is provided to predict drug-target interactions based on RLS classifier of Kronecker product kernel and GIP kernel similarities of drugs and targets [41]. SDTRLS is provided to predict drug-target interactions based on integration similarity of drug GIP kernel similarity and chemical substructure similarity by the SNF method[42, 47]. LDAP is used to predict lncRNA-disease associations by using a bagging SVM

Table 1 The description of benchmark dataset

Data type	Data	Database	dimensionality
chemical	Chemical substructures	PubChem	881
Biological	Drug-targets	DrugBank	780
	Drug transporters	DrugBank	18
	Drug enzymes	DrugBank	129
	Drug pathways	KEGG	253
Phenotypic	Drug indications	SIDER	4897
	Drug side effects	SIDER	4897
	Drug off side effects	OFFSIDES	9496
Interaction	Drug-drug interactions	TWOSIDES	Drugs:548,DDIs:48,584

Yan *et al. BMC Bioinformatics* 2019, **20**(Suppl 15):538

Page 4 of 10

classifier based on lncRNA and disease similarities which include GIP kernel similarity [43]. DNRLMF-MDA is an miRNA-disease associations prediction method based on dynamic neighborhood regularized logistic matrix factorization, which also uses the GIP kernel similarity.

We compute the GIP similarity of drugs via known DDIs in this study. We denote $D = \{d_1, d_2, \ldots, d_N\}$ as the set of N drugs. The known DDIs can be represented by an adjacency matrix $Y \in N * N$. The value of y_{ij} is 1 if d_i and d_j have a known interaction, and 0 otherwise. The GIP kernel similarity between drugs d_i and d_j can be calculated as follows:

$$G_{sim}\left(d_i, d_j\right) = exp\left(-\gamma_d ||yd_i - yd_j||^2\right) \qquad (1)$$

$$\gamma_d = \gamma'_d / \left(\frac{1}{N}\sum_{i=1}^{N} ||yd_i||^2\right) \qquad (2)$$

where γ_d is the regularization parameter of kernel bandwidth and γ'_d is set to be 1 according to previous studies [42, 44], $yd_i = \{y_{i1}, y_{i2}, \ldots, y_{iN}\}$ is the interaction profile of drug d_i.

RLS classifier and prediction dDIs

The (kernel) RLS classifier is based on the assumption that similar principal (adjuvant) drugs are tended to interact with the same adjuvant (principal) drug and has been widely used in other areas [42, 48, 49]. After calculating the GIP kernel similarity G_{sim}, we adopt the RLS classifier to compute the interaction probability scores of drug pairs as follows:

$$\hat{Y}_p = G_{sim}(G_{sim} + \sigma I)^{-1} Y \qquad (3)$$

$$Y_P = \frac{\hat{Y}_p + \hat{Y}_p^T}{2} \qquad (4)$$

where σ is the regularization parameter and set to be 1 according to previous study [41]. Furthermore, the G_{sim} and I are the GIP similarity matrix and the identity matrix, respectively. The Y_p is the final prediction result matrix, which is symmetric. The interacted probabilities of drug pairs are ranked in descending order. A candidate drug pair with the rank 1 is of the most possible drug pair.

KNN for new drugs

New drugs have no any known interaction with other drugs, which makes prediction DDIs for these drugs is impossible by existing methods. Therefore, we adopt the KNN method to calculate their initial relational scores based on the integrated feature similarity of chemical structure, biological and phenotypic information.

In order to calculate the integrated feature similarity $S_{sim} \in N * N$, we adopt the Pearson correlation coefficient to compute similarity based on the binary vectors of drug chemical substructures, drug targets, drug transporters, drug enzymes, drug pathways, drug indications, drug side

effects and drug off side effects. We can see from Table 1 that the total dimensionality of a binary vector of any drug is 21,351, whose value is 1 when the related feature is present and otherwise is 0. Specifically, the similarity of drug pair d_i and d_j is calculated as follows:

$$S_{sim}^{i,j} = \frac{Cov\left(v_{d_i}, v_{d_j}\right)}{D\left(v_{d_i}\right)D\left(v_{d_j}\right)} = \frac{E\left(\left(v_{d_i} - Ev_{d_i}\right)\left(v_{d_j} - Ev_{d_j}\right)\right)}{D\left(v_{d_i}\right)D(v_{d_j})} \qquad (5)$$

where v_{d_i} and v_{d_j} are the feature vectors of drugs d_i and d_j, respectively. Cov is the covariance. E and D are the mathematical expectation and standard deviation, respectively.

After obtaining the integrated feature similarity S_{sim}, we calculate the initial scores of new drugs by the KNN method. Specifically, the interaction scores $Y_{KNN}(d_i, d_j)$ between new drug d_i and another drug d_j can be calculated as follows:

$$Y_{KNN}\left(d_i, d_j\right) = \frac{\sum S_{sim}^{(i,l)} y_{lj}}{\sum S_{sim}^{(i,l)}}, d_l \in K_{set} \qquad (6)$$

where $S_{sim}^{(i,l)}$ is the (i, l)-th element of the integrated similarity matrix and y_{lj} is the (l, j)-th element of known DDIs matrix $Y \in N * N$. K_{set} represents the set of top K nearnest neighbors according to the S_{sim} matrix. In this study, we set the value of K by de novo drug validation.

Algorithm 1 is the description of our DDIGIP method. As the 0 vectors in the DDIs adjacency matrix Y correspond to unknown cases, we firstly compute the initial relational interaction scores for new drugs via the KNN method which uses the feature similarity S_{sim} of drugs by integrating chemical, biological and phenotypic data. The

---

Algorithm 1 DDIGIP

Input: Drug set DS, DDIs adjacency matrix Y, and parameter K

Output: predicted association matrix Y_p

DDIGIP(DS, Y, K)

1: Search for drug set D_{New}(0 vector) that each drug has no known DDIs;
2: if D_{New} is not null
3:    Calculate feature similarity S_{sim} of drugs by Pearson correlation coefficient;
4:    for i in D_{New}
5:       Calculate the $Y_{KNN}(i, :)$ by the KNN method;
6:       Set initial scores of drug d_i by $Y_{KNN}(i, :)$;
7:    end
8: end
9: Calculate GIP similarity G_d of drugs;
10: Based on G_d and Y, obtain the prediction result Y_p by the RLS method;
11: Return Y_p;

Yan *et al. BMC Bioinformatics* 2019, **20**(Suppl 15):538

Page 5 of 10

feature similarity S_{sim} is calculated by Pearson correlation coefficient. After computing the GIP similarity G_d of drugs, we take the RLS classifier to calculate the interaction scores of drug pairs. The final prediction result matrix is Y_p. Figure 1 demonstrates the work flow of DDIGIP.

RESULTS aND dISCUSSIONS

Benchmark evaluation and evaluation indices

5CV and 10CV are widely used to evaluate the performance for predicting drug-drug interactions [28, 29], drug-target interactions [42, 50], drug-disease interactions [51–53], lncRNA-disease associations [43, 54], miRNA-disease associations [44, 55] and so on. In this study, we evaluate the predictive performance of DDIGIP using 5CV and 10CV. In 5CV, all known DDIs are divided into 5 folds, and each fold, in turn, was left out as the test set while the rest 4 folds as the training set. In 10CV, we also divide known DDIs into 10 folds, and each fold is treated as test set in turn, while the remaining 9-folds are as the training set. We adopt 10 repeats in 5CV and 10CV. Furthermore, the actual generalization ability of predicting potential DDIs for new drugs is also an important aspect to assess the prediction performance. We thus conduct de novo drug validation to evaluate the predictive performance of DDIGIP. In de novo drug validation, we take known DDIs of each drug, in turn, and the rest DDIs of other drugs as the training set.

From a prediction method, each drug pair obtains a prediction score. Then each known interaction between two drugs in the test is ranked relative to the candidate interactions (all unknown interactions). On a specified rank threshold, TPR (true positive rate) is the fraction of known interactions that are correctly predicted, and FPR (false positive rate) is the fraction of unknown interactions that are predicted to be true interactions. The receiver operating characteristic curve (ROC) can be drawn based on various TPR and FPR values with various rank thresholds. Then we also use the area under the receiver operating curve (AUC) to measure the prediction performance of DDIGIP and other methods. The higher its AUC value is, the better prediction performance a method achieves.

Comparison with previous methods

In this study, we compare our method with other four methods: weighted average ensemble (WAE) method, L1 classifier ensemble (L1E) method, L2 classifier ensemble (L2E) method [29] and label propagation (LP) method [26], with the same validation method in the benchmark dataset.

5CV

Table 2 shows that the prediction performances of five methods in 5CV. Based on the AUC values of these methods, DDIGIP is slightly better than other methods. It shows that the GIP similarity is reasonable to use known DDIs because DDIGIP only uses known DDIs in 5CV. In addition, three integrating methods (WAE, L1E, L2E) were also achieved the good results because they integrated the neighbor recommender method, random walk method and matrix perturbation method.
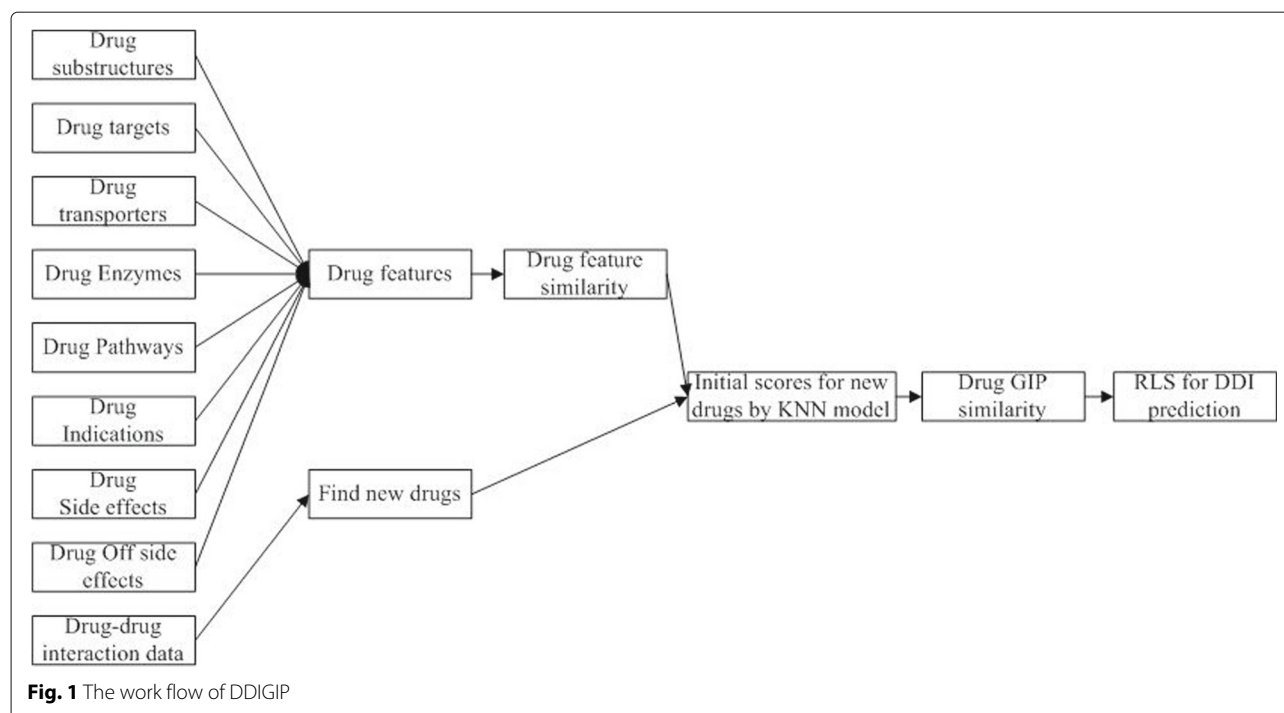

Fig. 1 The work flow of DDIGIP

Yan *et al. BMC Bioinformatics* 2019, **20**(Suppl 15):538

Page 6 of 10

Table 2 The prediction performances in 5CV,10CV and denovo validation, the best results are in the bold face

Method	Feature	5CV	10CV	Denovo
	The prediction performances(AUC)			
WAE	Chemical data, biological data, phenotypic data	0.9502	0.9530	0.9073
L1E	Chemical data, biological data, phenotypic data	0.9570	0.9599	Ø
L2E	Chemical data, biological data, phenotypic data	0.9561	0.9594	Ø
LP	Drug-sub	0.9356	0.9359	0.8993
	Drug-Label	0.9364	0.9368	0.8994
	Drug-Off Label	0.9374	0.9378	0.8997
DDIGIP	Chemical data, biological data, phenotypic data	**0.9600**	**0.9636**	**0.9262**

The Ø represents that we did not compute the prediction performance because the prediction limit for new drugs.

10CV

Table 2 also shows the prediction performances of five methods in 10CV. DDIGIP also achieved the best prediction result and its AUC value is 0.9636 which is larger than other methods WAE: 0.9530, L1E:0.9599, L2E:0.9594 and LP (max): 0.9378, respectively. By comparing the prediction performances of DDIGIP in 5CV and 10CV, DDIGIP is more effective to predict DDIs in 10CV than in 5CV. It proves that DDIGIP has better prediction ability when there are many known DDIs.

Denovo drug validation

In de novo drug validation, we compare DDIGIP with LP and WAE. We do not perform the de novo drug validation on other existing methods because of their prediction limit for new drugs. Similar to previous studies, we also obtain the weights of integrated methods (neighbor recommender method and random walk method) with drug chemical data, biological data and phenotypic data. Table 2 shows that DDIGIP also obtains the best prediction performance in terms of AUC (0.9262), compared with other methods (WAE: 0.9073, LP (max):0.8997). It also further indicates that the GIP similarity is effective to use known DDIs.

Computation time comparison

The computation time is also an important aspect to assess the performance of computational methods. In this study, we also compare the average computation time of five methods in 5CV. Figure 2 shows that the runtime of DDIGIP is less than those of other methods. In addition, since WAE, L1E and L2E are the integration

method, their computation times are longer than those of LP and DDIGIP. We can see from Fig. 2 that DDIGIP runs the fastest and its computation time is 6.61 seconds in 5CV.
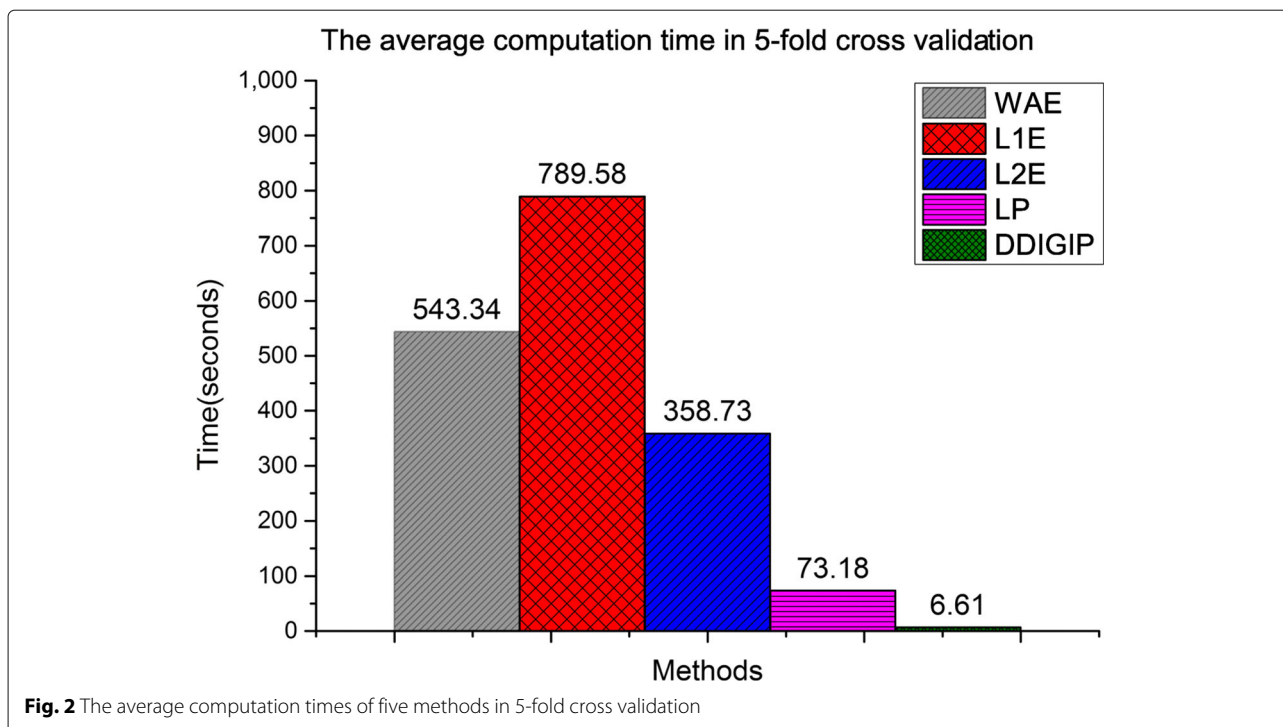
Parameter analysis for *K*

In order to verify the robustness of DDIGIP, we analyze the parameters *K* that is the number of the nearest neighbors in de novo drug validation. The optimal parameter value of *K* is selected by the grid search. Figure 3 shows the AUC values of DDIGIP under variation of K ranging from 1 to 15 in de novo validation. We can see from Fig. 3 that the prediction performance has the ascending trend when K ranges from 1 to 7, while has the descending trend when K ranges from 11 to 15. In addition, DDIGIP has a relatively stable prediction performance and achieves the best prediction result (AUC:0.9262) when K is 9. It indicates that a reasonable value of K can improve the prediction performance of DDIGIP.

Case studies

To illustrate the prediction performance of DDIGIP method, we conduct two types of case studies. The one includes the top 20 predicted DDIs under all known DDIs, in which the benchmark dataset is obtained from the TWOSIDES database while the confirmed database is DrugBanK database. Another includes top 20 the new DDIs in de novo validation of drug Ranolazine (DB00243) whose confirmed database composes of TWOSIDES database and DrugBanK database.

We can see from Table 3 that 9 out of top 20 DDIs predicted by DDIGIP are validated in DrugBank. The verification success rate is 45%. Zafirlukast (DB00549) is an oral leukotriene receptor antagonist (LTRA) drug usually used in the maintenance treatment of asthma, its metabolism can be decreased by Rabeprazole (DB01129) [56, 57]. Atazanavir (DB01072) is an antiretroviral drug of the protease inhibitor (PI) class, which is used to treat infection of human immunodeficiency virus (HIV) and its metabolism can be decreased when combining with Amlodipine (DB00381) [8, 58]. In addition, Pantoprazole (DB00213) also decreases the metabolism of Methadone (DB00333)[59]. The risk or severity of adverse effects can be increased when Atenolol (DB00335) is combined with Nadolol (DB01203), Clotrimazole (DB00257) is combined with Pregabalin (DB00230) or Enalapril (DB00584) is combined with Perindopril (DB00790) [9, 10, 60, 61]. The hypotensive activities of Nadolol (DB01203) can be increased by Propranolol (DB00571) [62]. The absorption of Cefpodoxime (DB01416) can be decreased when combining with Ranitidine (DB00863) [63]. Acebutolol (DB01193) also increases the serum concentration of Metoprolol (DB00264) [64].

Yan *et al. BMC Bioinformatics* 2019, **20**(Suppl 15):538

Page 7 of 10

Fig. 2 The average computation times of five methods in 5-fold cross validation

Ranolazine is an antianginal medication used in the treatment of chronic angina [10]. Table 4 shows that top 20 predicted DDIs of Ranolazine are validated in TWOSIDES database or DrugBanK database. In addition, 12 out of top 20 DDIs are simultaneously confirmed by TWOSIDES database and DrugBanK database, while the rest are confirmed by one of them. For example,

the metabolism of Levothyroxine (DB00451) and Zolpidem (DB00425) can be decreased when combining with Ranolazine [15, 56]. Clopidogrel is an antiplatelet agent structurally and pharmacologically similar to ticlopidine, which is used to inhibit blood clots in a variety of conditions such as peripheral vascular disease, coronary artery disease, and cerebrovascular disease [8].

Fig. 3 The AUC of DDIGIP under different settings of *K* in de novo drug validation, the sign * represents the default value

Yan *et al. BMC Bioinformatics* 2019, **20**(Suppl 15):538

Page 8 of 10
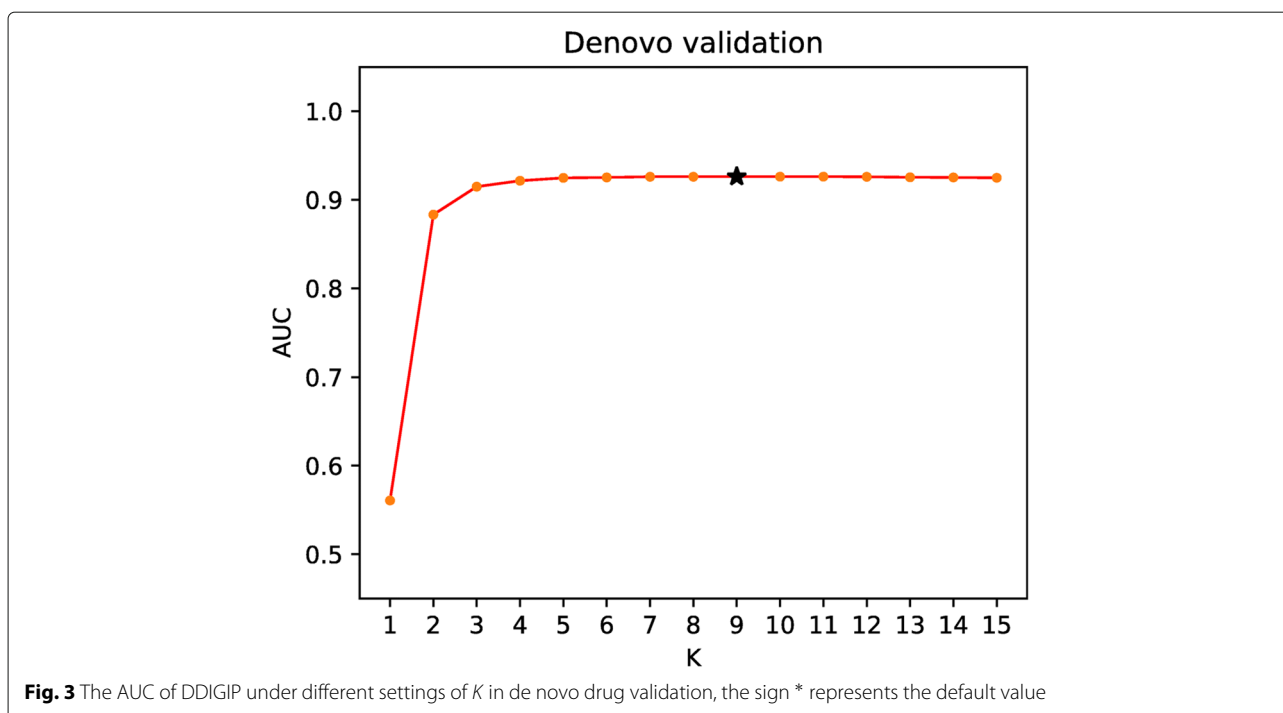
Table 3 Top 20 new DDIs predicted by DDIGIP method

Rank	Drug ID1	Drug ID2	Evidence
1	DB00448	DB01059	Unknown
2	DB00549	DB01129	DrugBank
3	DB00991	DB00231	Unknown
4	DB00470	DB00331	Unknown
5	DB00630	DB00346	Unknown
6	DB00863	DB01416	DrugBank
7	DB01203	DB00335	DrugBank
8	DB00813	DB00535	Unknown
9	DB00257	DB00230	DrugBank
10	DB00806	DB01036	Unknown
11	DB00927	DB01193	Unknown
12	DB00333	DB00213	DrugBank
13	DB00987	DB00758	Unknown
14	DB01595	DB01137	Unknown
15	DB06151	DB01068	Unknown
16	DB00328	DB00218	Unknown
17	DB00264	DB01193	DrugBank
18	DB01072	DB00381	DrugBank
19	DB01203	DB00571	DrugBank
20	DB00584	DB00790	DrugBank

Table 4 The validation result of top 20 new DDIs of drug Ranolazine (DB00243) predicted by DDIGIP method in de novo validation

Rank	Drug ID1	Drug ID2	Evidence
1	DB00243	DB00451	TWOSIDES,DrugBanK
2		DB00338	TWOSIDES,DrugBanK
3		DB00641	TWOSIDES,DrugBanK
4		DB00945	TWOSIDES,DrugBanK
5		DB00758	TWOSIDES,DrugBanK
6		DB00316	DrugBanK
7		DB00264	TWOSIDES
8		DB00695	TWOSIDES
9		DB00722	TWOSIDES
10		DB00390	TWOSIDES,DrugBanK
11		DB00448	TWOSIDES,DrugBanK
12		DB00999	TWOSIDES
13		DB00863	TWOSIDES,DrugBanK
14		DB00630	TWOSIDES
15		DB00635	DrugBanK
16		DB00213	TWOSIDES,DrugBanK
17		DB00678	TWOSIDES,DrugBanK
18		DB00425	TWOSIDES,DrugBanK
19		DB00177	TWOSIDES
20		DB00331	TWOSIDES,DrugBanK

The serum concentration of Clopidogrel (DB00758) can be increased when combining with Ranolazine [15]. Similarly, the serum concentration of Simvastatin (DB00641), Acetylsalicylic (DB00945) or Metformin (DB00331) also can be increased when combining with Ranolazine [56, 65]. In addition, when Ranolazine is combined with Omeprazole (DB00338) or Acetaminophen (DB00316), its serum concentration also can be increased [15, 66].

Conclusion

In this study, we have proposed a computational method, called DDIGIP, for DDIs prediction. The GIP similarity of drugs is calculated by the known DDIs, which makes full use of known DDIs. To our knowledge, in the previous studies the RLS-Kron method is used to predict interaction of bipartite networks, such as drug-target interaction networks, drug-disease interaction network and so on. Experiments are conducted using two different types of cross validations: 5-fold cross validation and 10-fold cross validation. The prediction ability of DDIGIP has been illustrated by comparing it with four other competing state-of-the-art methods.

Furthermore, based on Pearson correlation coefficient, we obtain a comprehensive feature similarity of drugs by integrating the chemical, biological and phenotypic data into a high dimension binary vector. In order to more effectively predict DDIs for new drugs, we also conduct de novo drug validation. We add a preprocessing step, KNN, to compute the initial relational scores according to the feature similarity of drugs. Because the vector 0 in the matrix corresponding to unknown cases or missing values rather than confirmed non-interactions, the preprocessing can improve the prediction performance.

Despite the advantages of DDIGIP as discussed above, it still has some limitations. The more effective method should be developed to integrate known DDIs with other chemical, biological and phenotypic data. In addition, other new prediction methods such as matrix completion [67], deep learning [68] and interpretable boosting model [69] could be considered. Finally, in this study, the benchmark dataset of DDIs only includes the positive samples and is an imbalanced dataset, we will also consider some other methods (SVM [70],LibD3C [71],extreme learning machine [72] and collaborative metric learning [73]) to predict DDIs when we obtain reliable negative samples in the future. We expect to develop a more effective method to predict DDIs by overcoming these limitations in the future.

Abbreviations
10CV: 10-fold cross validation; 5CV: 5-fold cross validation; AUC: Area under the receiver operating curve; DDI: Drug-drug interaction; GIP: Gaussian interaction profile; KNN: K-nearest neighbors; L1E: L1 Classifier ensemble method; L2E:L2 Classifier ensemble method; LP: Label propagation; RLS: Regularized least squares classifier; WAE: Weighted average ensemble method

Yan *et al. BMC Bioinformatics* 2019, **20**(Suppl 15):538

Page 9 of 10

Author details
[1] School of Computer Science and Engineering, Central South University, 932 South Lushan Rd, 410083, ChangSha, China. [2] School of Computer and Information,Qiannan Normal University for Nationalities, Longshan Road, 558000, DuYun, China. [3] Department of Computer Science, Georgia State University, GA30302, Atlanta, USA. [4] Biomedical Engineering and Department of Mechanical Engineering, University of Saskatchewan, SKS7N5A9, Saskatoon, Canada.

References
1. Crowther NR, Holbrook AM, Kenwright R, Kenwright M. Drug interactions among commonly used medications. chart simplifies data from critical literature review. Can Fam Phys. 1997;43:1972.
2. Venkatakrishnan K, von Moltke LL, Obach R, Greenblatt DJ. Drug metabolism and drug interactions: application and clinical value of in vitro models. Curr Drug Metabolism. 2003;4(5):423–59.
3. Quinn D, Day R. Drug interactions of clinical importance. Drug Safety. 1995;12(6):393–452.
4. Onakpoya IJ, Heneghan CJ, Aronson JK. Post-marketing withdrawal of anti-obesity medicinal products because of adverse drug reactions: a systematic review. BMC Med. 2016;14(1):191.
5. CDC. Health, United States, 2014 (5/2015)-hus14.pdf. http://www.cdc.gov/nchs/data/hus/hus14.pdf. Accessed 15 Nov 2017.
6. Nahta R, Hung M-C, Esteva FJ. The her-2-targeting antibodies trastuzumab and pertuzumab synergistically inhibit the survival of breast cancer cells. Cancer Res. 2004;64(7):2343–6.
7. Chou T-C. Drug combination studies and their synergy quantification using the chou-talalay method. Cancer Res. 2010;70(2):440–6.
8. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, Chang Z, Woolsey J. Drugbank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res. 2006;34(suppl_1): 668–72.
9. Knox C, Law V, Jewison T, Liu P, Ly S, Frolkis A, Pon A, Banco K, Mak C, Neveu V, et al. Drugbank 3.0: a comprehensive resource for 'omics' research on drugs. Nucleic Acids Res. 2010;39(suppl_1):1035–41.
10. Law V, Knox C, Djoumbou Y, Jewison T, Guo AC, Liu Y, Maciejewski A, Arndt D, Wilson M, Neveu V, et al. Drugbank 4.0: shedding new light on drug metabolism. Nucleic Acids Res. 2013;42(D1):1091–7.
11. Wang Y, Xiao J, Suzek TO, Zhang J, Wang J, Bryant SH. Pubchem: a public information system for analyzing bioactivities of small molecules. Nucleic Acids Res. 2009;37(suppl_2):623–33.
12. Kuhn M, Campillos M, Letunic I, Jensen LJ, Bork P. A side effect resource to capture phenotypic effects of drugs. Mole Syst Biol. 2010;6(1):343.
13. Tatonetti NP, Patrick PY, Daneshjou R, Altman RB. Data-driven prediction of drug effects and interactions. Sci Trans Med. 2012;4(125):125–31.
14. Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M. Kegg for representation and analysis of molecular networks involving diseases and drugs. Nucleic Acids Res. 2009;38(suppl_1):355–60.
15. Poluzzi E, Raschi E, Moretti U, De Ponti F. Drug-induced torsades de pointes: data mining of the public version of the fda adverse event reporting system (aers). Pharmacoepidemiol Drug Safety. 2009;18(6):512–8.
16. Vilar S, Harpaz R, Uriarte E, Santana L, Rabadan R, Friedman C. Drug-drug interaction through molecular structure similarity analysis. J Am Med Inf Assoc. 2012;19(6):1066–74.
17. Liu L, Chen L, Zhang Y.-H., Wei L, Cheng S, Kong X, Zheng M, Huang T, Cai Y-D. Analysis and prediction of drug–drug interaction by minimum redundancy maximum relevance and incremental feature selection. J Biomole Struct Dynamics. 2017;35(2):312–29.
18. Cheng F, Zhao Z. Machine learning-based prediction of drug–drug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties. J Am Med Inf Assoc. 2014;21(e2):278–86.
19. Vilar S, Uriarte E, Santana L, Tatonetti NP, Friedman C. Detection of drug-drug interactions by modeling interaction profile fingerprints. PloS one. 2013;8(3):58321.
20. Takeda T, Hao M, Cheng T, Bryant SH, Wang Y. Predicting drug–drug interactions through drug structural similarities and interaction networks incorporating pharmacokinetics and pharmacodynamics knowledge. J Cheminforma. 2017;9(1):16.
21. Vilar S, Uriarte E, Santana L, Friedman C, P Tatonetti N. State of the art and development of a drug-drug interaction large scale predictor based on 3d pharmacophoric similarity. Curr Drug Metabolism. 2014;15(5):490–501.
22. Herrero-Zazo M, Segura-Bedmar I, Hastings J, Martínez P. Dinto: using owl ontologies and swrl rules to infer drug–drug interactions and their mechanisms. J Chem Inf Modeling. 2015;55(8):1698–707.
23. Li P, Huang C, Fu Y, Wang J, Wu Z, Ru J, Zheng C, Guo Z, Chen X, Zhou W, et al. Large-scale exploration and analysis of drug combinations. Bioinformatics. 2015;31(12):2007–16.
24. Ferdousi R, Safdari R, Omidi Y. Computational prediction of drug-drug interactions based on drugs functional similarities. J Biomed Informa. 2017;70:54–64.
25. Sridhar D, Fakhraei S, Getoor L. A probabilistic approach for collective similarity-based drug–drug interaction prediction. Bioinformatics. 2016;32(20):3175–82.
26. Zhang P, Wang F, Hu J, Sorrentino R. Label propagation prediction of drug-drug interactions based on clinical side effects. Sci Rep. 2015;5:12339.
27. Cao D-S, Xiao N, Li Y-J, Zeng W-B, Liang Y-Z, Lu A-P, Xu Q-S, Chen A. Integrating multiple evidence sources to predict adverse drug reactions based on a systems pharmacology model. CPT: Pharmacom Syst Pharmacol. 2015;4(9):498–506.
28. 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(7):1004975.
29. Zhang W, Chen Y, Liu F, Luo F, Tian G, Li X. Predicting potential drug-drug interactions by integrating chemical, biological, phenotypic and network data. BMC bioinformatics. 2017;18(1):18.
30. Matthews EJ, Frid AA. Prediction of drug-related cardiac adverse effects in humans-a: Creation of a database of effects and identification of factors affecting their occurrence. Reg Toxicol Pharmacol. 2010;56(3):247–75.

Yan *et al. BMC Bioinformatics* 2019, **20**(Suppl 15):538

Page 10 of 10

31. Zakharov AV, Varlamova EV, Lagunin AA, Dmitriev AV, Muratov EN, Fourches D, Kuz'min VE, Poroikov VV, Tropsha A, Nicklaus MC. Qsar modeling and prediction of drug–drug interactions. Mole Pharma. 2016;13(2):545–56.

32. Cami A, Manzi S, Arnold A, Reis BY. Pharmacointeraction network models predict unknown drug-drug interactions. PloS one. 2013;8(4):61468.

33. Huang H, Zhang P, Qu XA, Sanseau P, Yang L. Systematic prediction of drug combinations based on clinical side-effects. Sci Rep. 2014;4:7160.

34. Tari L, Anwar S, Liang S, Cai J, Baral C. Discovering drug–drug interactions: a text-mining and reasoning approach based on properties of drug metabolism. Bioinformatics. 2010;26(18):547–53.

35. Iyer SV, Harpaz R, LePendu P, Bauer-Mehren A, Shah NH. Mining clinical text for signals of adverse drug-drug interactions. J Am Med Informa Assoc. 2014;21(2):353–62.

36. Banda JM, Callahan A, Winnenburg R, Strasberg HR, Cami A, Reis BY, Vilar S, Hripcsak G, Dumontier M, Shah NH. Feasibility of prioritizing drug–drug-event associations found in electronic health records. Drug Safety. 2016;39(1):45–57.

37. Chen L, Chu C, Zhang Y-H, Zheng M, Zhu L, Kong X, Huang T. Identification of drug-drug interactions using chemical interactions. Curr Bioinforma. 2017;12(6):526–34.

38. Wang J, Kribelbauer J, Rabadan R. Network propagation reveals novel features predicting drug response of cancer cell lines. Curr Bioinforma. 2016;11(2):203–10.

39. Liu H, Luo L, Cheng Z, Sun J, Guan J, Zheng J, Zhou S. Group-sparse modeling drug-kinase networks for predicting combinatorial drug sensitivity in cancer cells. Curr Bioinforma. 2018;13(5):437–43.

40. Öztürk H, Özgür A, Ozkirimli E. Deepdta: deep drug–target binding affinity prediction. Bioinformatics. 2018;34(17):821–9.

41. van Laarhoven T, Nabuurs SB, Marchiori E. Gaussian interaction profile kernels for predicting drug–target interaction. Bioinformatics. 2011;27(21):3036–43.

42. Yan C, Wang J, Lan W, Wu F-X, Pan Y. Sdtrls: Predicting drug-target interactions for complex diseases based on chemical substructures. Complexity. 2017;2017(Article ID 2713280):10.

43. Lan W, Li M, Zhao K, Liu J, Wu F-X, Pan Y, Wang J. Ldap: a web server for lncrna-disease association prediction. Bioinformatics. 2016;33(3):458–60.

44. Yan C, Wang J, Ni P, Lan W, Wu FX, Pan Y. Dnrlmf-mda:predicting microrna-disease associations based on similarities of micrornas and diseases. IEEE/ACM Trans Comput Biol Bioinforma. 2019;16(1):233–43.

45. Yan C, Duan G, Wu F, Pan Y, Wang J. Brwmda: Predicting microbe-disease associations based on similarities and bi-random walk on disease and microbe networks. IEEE/ACM Trans Comput Biol Bioinforma. 2019. https://doi.org/10.1109/TCBB.2019.2907626. (to be published).

46. Yan C, Duan G, Wu F, Pan Y, Wang J. Mchmda: Predicting microbe-disease associations based on similarities and low-rank matrix completion. IEEE/ACM Trans Comput Biol Bioinforma. 2019. https://doi.org/10.1109/TCBB.2019.2926716. (to be published).

47. Wang B, Mezlini AM, Demir F, Fiume M, Tu Z, Brudno M, Haibe-Kains B, Goldenberg A. Similarity network fusion for aggregating data types on a genomic scale. Nat Methods. 2014;11(3):333.

48. Xia Z, Wu L-Y, Zhou X, Wong ST. Semi-supervised drug-protein interaction prediction from heterogeneous biological spaces. BMC Syst Biol. 2010;4(2):S6.

49. Yan C, Wang J, Wu F-X. Dwnn-rls: regularized least squares method for predicting circrna-disease associations. BMC bioinformatics. 2018;19(19):520.

50. Wu Z, Cheng F, Li J, Li W, Liu G, Tang Y. Sdtnbi: an integrated network and chemoinformatics tool for systematic prediction of drug–target interactions and drug repositioning. Brief Bioinforma. 2016;18(2):333–47.

51. Luo H, Wang J, Li M, Luo J, Peng X, Wu F-X, Pan Y. Drug repositioning based on comprehensive similarity measures and bi-random walk algorithm. Bioinformatics. 2016;32(17):2664–71.

52. Luo H, Li M, Wang S, Liu Q, Li Y, Wang J. Computational drug repositioning using low-rank matrix approximation and randomized algorithms. Bioinformatics. 2018;34(11):1904–12.

53. Luo H, Wang J, Li M, Luo J, Ni P, Zhao K, Wu FX, Pan Y. Computational drug repositioning with random walk on a heterogeneous network. IEEE/ACM Trans Comput Biol Bioinforma. 2018. https://doi.org/10.1109/TCBB.2018.2832078. (to be published).

54. Lu C, Yang M, Luo F, Wu F-X, Li M, Pan Y, Li Y, Wang J. Prediction of lncrna-disease associations based on inductive matrix completion. Bioinformatics. 2018;34(19):3357–64.

55. Lan W, Wang J, Li M, Liu J, Wu F-X, Pan Y. Predicting microrna-disease associations based on improved microrna and disease similarities. IEEE/ACM Trans Comput Biol Bioinforma. 2018;15(6):1774–82.

56. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, Sajed T, Johnson D, Li C, Sayeeda Z, et al. Drugbank 5.0: a major update to the drugbank database for 2018. Nucleic Acids Res. 2017;46(D1):1074–82.

57. Bhattacharya A, Bandichhor R. Green technologies in the generic pharmaceutical industry. Green Chem Pharma Ind. 2010304–6.

58. Mauss S, Klinker H. Drug-drug interactions in the treatment of hcv among people who inject drugs. Clin Inf diseases. 2013;57(suppl_2):125–8.

59. Welage LS, Berardi RR. Evaluation of omeprazole, lansoprazole, pantoprazole, and rabeprazole in the treatment of acid-related diseases. J Am Pharma Assoc 1996. 2000;40(1):52–62.

60. Vítovec J, Špinar J. First-dose hypotension after angiotensin-converting enzyme (ace) inhibitors in chronic heart failure: a comparison of enalapril and perindopril. Eur J Heart Fail. 2000;2(3):299–304.

61. Burmeister WE, Reynolds RD, Lee RJ. Limitation of myocardial infarct size by atenolol, nadolol and propranolol in dogs. Eur J Pharmacol. 1981;75(1):7–10.

62. Reeves RA, From GL, Paul W, Leenen FH. Nadolol, propranolol, and thyroid hormones: Evidence for a membrane-stabilizing action of propranolol. Clin Pharmacol Therap. 1985;37(2):157–61.

63. UCHIDA E, OGUCHI K, HISAOKA M, KOBAYASHI S, KAI K, YASUHARA H. Effects of ranitidine, metoclopromide, and anisotropine methylbromide on the availability of cefpodoxime proxetil (cs-807) in japanese healthy subjects. Rinsho yakuri/Japan J Clin Pharmacol Therap. 1988;19(3):573–9.

64. El-Beqqali A, Kussak A, Blomberg L, Abdel-Rehim M. Microextraction in packed syringe/liquid chromatography/electrospray tandem mass spectrometry for quantification of acebutolol and metoprolol in human plasma and urine samples. J Liquid Chromatogr Rel Technol. 2007;30(4):575–86.

65. Florentin M, Elisaf MS. Simvastatin interactions with other drugs. Exp Opin Drug Safety. 2012;11(3):439–44.

66. Treyger G, Silver SA, Sakharova AA. Pheochromocytoma diagnosis after an abnormal stress test: Case report and review of the literature. J. 2015;115(7):3.

67. Yang M, Luo H, Li Y, Wang J. Drug repositioning based on bounded nuclear norm regularization. Bioinformatics. 2019;35(14):i455–63. https://doi.org/10.1093/bioinformatics/btz331. (to be published).

68. Liu J, Pan Y, Li M, Chen Z, Tang L, Lu C, Wang J. Applications of deep learning to mri images: A survey. Big Data Mining Anal. 2018;1(1):1–18.

69. Liu L, Yu Y, Fei Z, Li M, Wu F-X, Li H-D, Pan Y, Wang J. An interpretable boosting model to predict side effects of analgesics for osteoarthritis. BMC Syst Biol. 2018;12(6):105.

70. Chang C-C, Lin C-J. Libsvm: a library for support vector machines. ACM Trans Intell Syst Technol (TIST). 2011;2(3):27.

71. Lin C, Chen W, Qiu C, Wu Y, Krishnan S, Zou Q. Libd3c: ensemble classifiers with a clustering and dynamic selection strategy. Neurocomputing. 2014;123:424–35.

72. Wang H, Wang J, Zhou L. A survival ensemble of extreme learning machine. Appl Intell. 2018;48(7):1846–58.

73. Luo H, Wang J, Yan C, Li M, Fangxiang W, Yi P. A novel drug repositioning approach based on collaborative metric learning. IEEE/ACM Trans Comput Biol Bioinforma. 2019. https://doi.org/10.1109/TCBB.2019.2926453. (to be published).

Publisher's Note