Bioinformatics, 36(9), 2020, 2839–2847
doi: 10.1093/bioinformatics/btaa062
Advance Access Publication Date: 30 January 2020
Original Paper

Systems biology

DRIMC: an improved drug repositioning approach using Bayesian inductive matrix completion

Wenjuan Zhang[1], Hunan Xu[1], Xiaozhong Li[2], Qiang Gao[3] and Lin Wang[2,*]

[1]College of General Education, Tianjin Foreign Studies University, Tianjin 300204, China, [2]College of Artificial Intelligence, Tianjin University of Science and Technology, Tianjin 300457, China and [3]Key Lab of Industrial Fermentation Microbiology, Ministry of Education & Tianjin City, College of Biotechnology, Tianjin University of Science and Technology, Tianjin 300457, China

*To whom correspondence should be addressed.

Associate Editor: Anthony Mathelier

Received on September 9, 2019; revised on January 20, 2020; editorial decision on January 21, 2020; accepted on January 22, 2020

Abstract

Motivation: One of the most important problems in drug discovery research is to precisely predict a new indication for an existing drug, i.e. drug repositioning. Recent recommendation system-based methods have tackled this problem using matrix completion models. The models identify latent factors contributing to known drug-disease associations, and then infer novel drug-disease associations by the correlations between latent factors. However, these models have not fully considered the various drug data sources and the sparsity of the drug-disease association matrix. In addition, using the global structure of the drug-disease association data may introduce noise, and consequently limit the prediction power.

Results: In this work, we propose a novel drug repositioning approach by using Bayesian inductive matrix completion (DRIMC). First, we embed four drug data sources into a drug similarity matrix and two disease data sources in a disease similarity matrix. Then, for each drug or disease, its feature is described by similarity values between it and its nearest neighbors, and these features for drugs and diseases are mapped onto a shared latent space. We model the association probability for each drug-disease pair by inductive matrix completion, where the properties of drugs and diseases are represented by projections of drugs and diseases, respectively. As the known drug-disease associations have been manually verified, they are more trustworthy and important than the unknown pairs. We assign higher confidence levels to known association pairs compared with unknown pairs. We perform comprehensive experiments on three benchmark datasets, and DRIMC improves prediction accuracy compared with six stat-of-the-art approaches.

Availability and implementation: Source code and datasets are available at https://github.com/linwang1982/DRIMC.

Contact: linwang@tust.edu.cn

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

The discovery of a new drug is a risky, laborious and costly process (Paul *et al.*, 2010). While the investments of drug development are rising, the number of approved drugs per year remains low. The last decades have witnessed that around 30% of new drug failures are due to safety issues found in clinical trials (Persidis *et al.*, 2011). Drug repositioning refers to rediscovering a new indication for an existing drug. Since repositioned drugs have already passed safety tests in clinical trials, these de-risked compounds can facilitate the process of drug discovery with lower overall development costs and shorter development timelines (Pushpakom *et al.*, 2019). Though the most successful examples of drug repositioning have been obtained through serendipitous or rational observations, computational drug repositioning methods can facilitate the development of drug repositioning, for which the identified ranked lists of candidate indications for existing drugs can be used to guide time consuming and costly wet experiments.

So far, various methods have been developed for the drug-disease association prediction, and can be roughly categorized into three groups. The first group includes machine learning-based methods by using known drug-disease associations, drug features and disease features. The logistic regression, Naïve Bayes, support vector machine and random forest were exploited to perform drug repositioning prediction, respectively (Gottlieb *et al.*, 2011; Oh *et al.*, 2014; Wang *et al.*, 2013; Yang and Agarwal, 2011). With the creation of various biological data, several heterogeneous networks were built. The second group covers network-based methods that try to capture missing drug-disease edges and their reliability on the heterogeneous networks. A model of triple layer heterogeneous

graph-based inference predicted drug-disease associations and drug-target interactions simultaneously (Wang *et al.*, 2014b). The method built a heterogeneous network, composed of information on drugs, diseases and targets, and performed computational drug repositioning using iterative algorithm that propagates information across the three-layer graph. Network-based prioritization method (DrugNet) simultaneously integrates information on diseases, drugs and targets to generate a ranked list of new candidate indications for a given drug query and vice versa (Martínez *et al.*, 2015). A method of utilizing comprehensive similarity measures and bi-random walk algorithm (MBiRW) was applied on the drug-disease heterogeneous network to perform drug repositioning (Luo *et al.*, 2016).

The third group refers to the matrix completion-based methods which aim to find the lowest rank matrix or a matrix of rank r that matches the known drug-disease entries. Drug Repositioning Recommendation System (DRRS) constructed a large drug-disease adjacency matrix, including entries for drug pairs, disease pairs, known drug-disease associations and unknown drug-disease associations, and adopted singular value thresholding algorithm for drug repositioning (Luo *et al.*, 2018). Recent approaches such as Laplacian regularized sparse subspace learning (LRSSL), non-negative matrix factorization (DisDrugPred) and similarity constrained matrix factorization (SCMFDD) were also presented to identify candidate therapeutic indications for drugs, respectively (Liang *et al.*, 2017; Wang *et al.*, 2017; Xuan *et al.*, 2019; Zhang *et al.*, 2018). However, these approaches have not fully utilized the various drug data sources and the sparsity of the drug-disease association matrix. Besides, these methods expressed drug and disease side information as similarity matrices, and generally the usage of the global structure of the drug-disease association data, i.e. considering all similar neighbors, may introduce noise and thus lower the prediction accuracy.

In this work, we propose a novel computational drug repositioning method by using Bayesian inductive matrix completion (DRIMC). Figure 1 shows the flowchart which illustrates the whole procedure in the proposed DRIMC method. First, we integrate four drug data sources, i.e. drug chemical structure, Pfam domain annotation of drug targets, gene ontology term of targets and drug-disease profile information, into a fused drug similarity matrix. Likewise we obtain a fused disease similarity matrix on the basis of the integration of disease phenotype and drug-disease profile information. Then, for each drug or disease, its feature is described by similarity values between it and its k-nearest neighbors, and these features for drugs and diseases are projected into a common subspace. The DRIMC method focuses on modeling the probability that a drug would associate with a disease by inductive matrix completion (IMC), where the properties of drugs and diseases are represented by projections of drugs and diseases, respectively. To demonstrate its effectiveness, we applied DRIMC to three benchmark datasets and compared it with six state-of-the-art methods, using five trails of 10-fold cross-validation (CV). The results showed that the prediction accuracy of DRIMC significantly exceeded the other methods. Moreover, we applied DRIMC to LRSSL dataset to perform computational drug repositioning, i.e. all already available drug-disease associations were used as the training set, and then prioritized the unknown associations based on the prediction scores. Finally, we compiled an independent test set to perform *de novo* drug-disease association prediction.

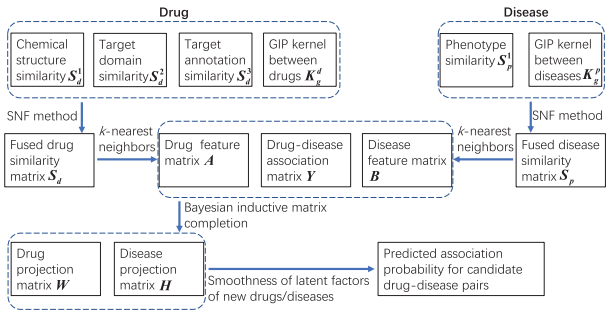

Fig. 1. Flowchart of the computational drug repositioning method DRIMC

2 Materials and methods

2.1 Data and preprocessing

Three datasets, i.e. PREDICT, Cdataset and LRSSL (Gottlieb *et al.*, 2011; Luo *et al.*, 2016; Liang *et al.*, 2017), were used to validate the effectiveness of our drug-disease association prediction method. The PREDICT dataset consists of 593 drugs, 313 diseases and 1933 validated drug-disease associations totally. The Cdataset includes 2353 known drug-disease associations involving 658 drugs and 409 diseases. The last dataset LRSSL includes 763 drugs, 681 diseases and 3051 known drug-disease associations. For each dataset, the data sources for drugs include drug chemical structure, Pfam domain annotation of drug targets and gene ontology term of targets, and were imported from DrugBank (David *et al.*, 2018, version 5.1.2). Diseases were extracted from Online Mendelian Inheritance in Man (OMIM) database (Hamosh *et al.*, 2002). Here we represent the associations between drugs and diseases by a binary matrix $Y \in R^{m\times n}$, where each entry $y_{ij} \in \{0,1\}$, m and n are the number of drugs and diseases in the dataset, respectively. We set y_{ij} =1 if the drug d_i has been experimentally validated to associate with the disease p_j, otherwise it is unknown and $y_{ij} = 0$.

Based on the chemical structure of drugs, PubChem fingerprint descriptors were computed using the PaDEL software (Yap, 2011). For each pair of drugs, the similarities between their feature profiles such as fingerprint descriptors, domain annotations and gene ontology terms were measured by the Jaccard coefficient, respectively. These drug similarities are represented by three $m \times m$ matrices $S_d^b (b = 1, 2, 3)$, where the (i, μ) entry $S_d^b(i, \mu)$ is the similarity between d_i and d_μ in the bth view. The disease similarity was calculated using MimMiner (Van Driel *et al.*, 2006) which is a text-mining approach to quantify relationships between human disease phenotypes from the OMIM database. This disease similarity is described using an $n \times n$ matrix S_p^1, where the (j, ν) entry $S_p^1(j, \nu)$ is the similarity between p_j and p_ν.

Furthermore, to fully utilize the topology of drug-disease association network, we computed the fourth drug similarity matrix, i.e. Gaussian interaction profile (GIP) kernel between drugs (denoted by K_g^d). Since new drugs had no association information in the training process, a reformulated drug-disease association matrix Y_r was first calculated in the following manner: initializing $Y_r = Y$ and for each new drug d_i, replacing the ith row of Y_r with new one inferred according to the following formula: $Y_r(i) = \frac{1}{\sum_{d_w \in N^+(d_i)} S_d^1(i,w)} \sum_{d_w \in N^+(d_i)} S_d^1(i,w) Y(w)$, where $Y_r(i)$ is the ith row of Y_r, $N^+(d_i)$ denotes the set of neighbors of d_i composing of k (empirically set to 20) known drugs, $Y(w)$ is the wth row of Y. Then, K_g^d was calculated by: $K_g^d(i,j) = \exp\left(-\frac{\|Y_r(i) - Y_r(j)\|^2}{\sigma_1}\right)$, where $Yr(j)$ is the jth row of Yr, $\|.\|$ denotes the Euclidean distance, and σ_1 is the kernel bandwidth which is empirically set as the average association number for each drug as described in previous work (Hao *et al.*, 2016). It should be noted that K_g^d had to be recalculated since the adjacency matrix Y changed when performing CV prediction. Likewise, another reformulated drug-disease association matrix Y_c was constructed, in which the jth column of Y_c (corresponding to the new disease p_j) was replaced in a similar way. Then, the second disease similarity matrix, i.e. the GIP kernel between diseases (denoted by K_g^p), was computed by using Y_c. Notably, the reformulation operation was only for creating profiles for GIP kernel construction, and did not modify the response matrix Y actually.

In summary, there are four drug similarity matrices (i.e. S_d^1, S_d^2, S_d^3 and K_g^d) and two disease similarity matrices (i.e. S_p^1 and K_g^p). We applied the similarity network fusion (SNF) method (Wang *et al.*, 2014a) to integrate four drug similarity matrices into a single fused drug similarity matrix S_d and two disease similarity matrices in a single fused disease similarity matrix S_p. The SNF is a non-linear

method based on message-passing theory, and can capture common as well as complementary information across different similarity measures. It iteratively updates every similarity network with information from the other networks, using nearest neighbors, making it more similar to the others. The details of SNF are illustrated in the Supplementary Material.

Furthermore, for each drug d_i, its feature is described by similarity values between d_i and its $k1$-nearest neighbors, and we can obtain drug feature matrix A whose elements $a_{i\mu}$ is defined as

$$a_{i\mu} = \begin{cases} S_d(i,\mu) & \text{if } d_\mu \in N(d_i) \\ 0 & \text{otherwise} \end{cases}, \tag{1}$$

where S_d is the drug similarity matrix and $N(d_i)$ is a set of drug d_i's neighbors including d_i in S_d. Similarly, we can obtain disease feature matrix B whose elements b_{jv} is defined as following:

$$b_{jv} = \begin{cases} S_p(j,v) & \text{if } p_v \in N(p_j) \\ 0 & \text{otherwise} \end{cases}, \tag{2}$$

where S_p is the disease similarity matrix and $N(p_j)$ is a set of disease p_j's neighbors.

2.2 Bayesian IMC

The IMC is adapted to recommender systems with side information of users and items. The primary idea of IMC is to map user feature space and item feature space onto a shared latent space through projection matrices W and H, respectively, and then find the optimal W and H such that the projections of users are geometrically close to the projections of their known associated items (Jain and Dhillon, 2013). The IMC model has been successfully applied to gene-disease association prediction (Natarajan and Dhillon, 2014) and miRNA-disease association prediction (Chen *et al.*, 2018). Here, we incorporate features associated with drugs and diseases in matrix completion, so that it enables prediction for drug repositioning.

We present a Bayesian treatment of IMC, where the logistic matrix factorization framework is adopted for solving W and H. Logistic matrix factorization was originally developed for recommender system (Johnson, 2014), and lately its variations were used for drug target prediction such as neighborhood regularized logistic matrix factorization (Liu *et al.*, 2016) and dual-network integrated logistic matrix factorization (Hao *et al.*, 2017). For drug d_i and disease p_j, the association probability $\Pr_{ij}$ is modeled by their projections with the following logistic function:

$$\Pr_{ij} = \frac{\exp\left(a_i W H^T b_j^T\right)}{1+\exp\left(a_i W H^T b_j^T\right)}, \tag{3}$$

where a_i is the ith row in A and b_j is the jth row in B.

As the known drug-disease associations have been manually verified, they are more trustworthy and important for improving prediction performance. Here, we define our confidence on the non-zero entries $y_{ij} = 1$ as αy_{ij}, where α is a tuning parameter and empirically set to 10. Thus, we represent each non-zero element $y_{ij} = 1$ as α positive observations and each zero element $y_{ij} = 0$ as a single negative observation. By making the assumption that all entries of Y are independent, we obtain the probability of the observations as follows:

$$\Pr(Y|W,H) = \prod_{i=1}^{m}\prod_{j=1}^{n} \Pr_{ij}^{\alpha y_{ij}} \left(1-\Pr_{ij}\right)^{\left(1-y_{ij}\right)}. \tag{4}$$

In addition, we place zero-mean spherical Gaussian priors on W and H to introduce regularization term for preventing overfitting:

$$\Pr\left(W|\sigma^2\right) = \prod_{i=1}^{m} N\left(w_i|0,\sigma^2 I\right), \tag{5}$$

$$\Pr\left(H|\sigma^2\right) = \prod_{j=1}^{n} N\left(h_j|0,\sigma^2 I\right), \tag{6}$$

where σ is the standard deviation of Gaussian distributions, w_i is the ith row in W, h_j is the jth row in H and I is the identity matrix. By using Bayesian inference, we have

$$\Pr(W,H|Y) \propto \Pr(Y|W,H)\Pr(W)\Pr(H). \tag{7}$$

Consequently, taking the ln of the posterior distribution we derive the following ln-posterior function,

$$\begin{aligned} \ln p\left(W,H|Y,\sigma^2\right) = &\sum_{i=1}^{m}\sum_{j=1}^{n}\Big\{\alpha y_{ij} a_i W H^T b_j^T \\ &-\left(1+\alpha y_{ij}-y_{ij}\right)\ln\left[1+\exp\left(a_i W H^T b_j^T\right)\right]\Big\} \\ &-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{m}||w_i||_2^2+\sum_{j=1}^{n}||h_j||_2^2\right)+C, \end{aligned} \tag{8}$$

where C is constant and independent of the model parameters (i.e. W and H).

2.3 Drug repositioning approach DRIMC

The model parameters W and H can then be learned by maximum a posterior (MAP) estimation

$$\begin{aligned} \max_{W,H} LP(W,H) = &\sum_{i=1}^{m}\sum_{j=1}^{n}\Big[\alpha Y \circ \left(A W H^T B^T\right) \\ &-(1+\alpha Y - Y)\circ \ln\left(1+\exp\left(A W H^T B^T\right)\right)\Big]_{ij} \\ &-\frac{\lambda}{2}||W||_F^2-\frac{\lambda}{2}||H||_F^2, \end{aligned} \tag{9}$$

where 1 is a matrix all of whose elements are one, $\circ$ denotes the Hadamard product of two matrices, $\lambda = \frac{1}{\sigma^2}$.

The alternating gradient ascend algorithm is used to solve for W and H from the above objective function. The partial gradients with respect to W and H are as follows:

$$\frac{\partial LP}{\partial W} = \alpha A^T Y B H - A^T Q B H - \lambda W, \tag{10}$$

$$\frac{\partial LP}{\partial H} = \alpha B^T Y^T A W - B^T Q^T A W - \lambda H, \tag{11}$$

where $Q = (1+\alpha Y - Y)\circ \frac{\exp\left(A W H^T B^T\right)}{1+\exp\left(A W H^T B^T\right)}$. To accelerate the convergence of the gradient ascent method, we choose the gradient step size adaptively via AdaGrad algorithm, where base learning rate η is determined by CV (Duchi *et al.*, 2011). We provide the details of the optimization algorithm for solving DRIMC model as follows.

Algorithm 1 DRIMC

Input: Drug-disease association matrix, $Y \in R^{m\times n}$; Drug chemical similarity matrix, $S_d^1 \in R^{m\times m}$; Drug target domain similarity matrix, $S_d^2 \in R^{m\times m}$; Drug target annotation similarity matrix, $S_d^3 \in R^{m\times m}$; Disease semantic similarity matrix, $S_p^1 \in R^{n\times n}$;
Output: Projection matrices, $W \in R^{m\times r}$ and $H \in R^{n\times r}$;
Step 1. Initialize W and H randomly with Gaussian distribution, where standard deviation $\sigma = \frac{1}{\sqrt{r}}$;
Step 2. Compute GIP kernel between drugs (K_g^d) and GIP kernel between diseases (K_g^p);
Step 3. Integrate S_d^1, S_d^2, S_d^3, K_g^d into S_d, and S_p^1, K_g^p in S_p by using SNF method;

Step 4. Construct drug feature matrix A and disease feature matrix B according to Equations (1) and (2), respectively;
Step 5. For $t = 1, \ldots, \max_{\text{iter}}$
Update W as follows:

$$W^t = W^{t-1} + \eta g_w^{t-1} \oslash \sqrt{\sum_{\tau=1}^{t-1} g_w^\tau \circ g_w^\tau}, \quad (12)$$

where g_w^τ denotes the partial gradient with respect to W at iteration τ which is computed according to Equation (10), and $\oslash$ is the Hadamard division.
Update H as follows:

$$H^t = H^{t-1} + \eta g_h^{t-1} \oslash \sqrt{\sum_{\tau=1}^{t-1} g_h^\tau \circ g_h^\tau}, \quad (13)$$

where g_h^τ denotes the partial gradient with respect to H at iteration τ which is computed according to Equation (11).

Step 6. Output W and H.

Besides, we can consider DRIMC model in another way. The model parameters W and H can be explained as drug and disease latent matrices, where each row (latent factor) corresponds to one drug or disease. Then, we represent each drug (d_i) or disease (p_j) by the linear combination of its neighbors' latent factors (a_{iW} or b_{jH}). Here, the set of neighbors of d_i (p_j) includes d_i (p_j), and the coefficient for d_i (p_j) is the largest (>0.5, while the coefficients for other neighbors are generally <0.2). If drug d_i associates with disease p_j, our model will necessitate their representation factors a_{iW} and b_{jH} geometrically close. After solving for the drug and disease latent matrices (W and H) from DRIMC model, the latent factor for each new drug or disease is not accurate enough, and thus the representation factor is not accurate. We smooth new drug/disease prediction by incorporating neighbor information as reported in the work of Liu *et al.* (2016). Specifically, for the new drug d_i, the ith row of W is replaced with new one inferred by using its $k2$-nearest neighbors according to the following formula:

$$\hat{w}_i = \frac{1}{\sum_{d_w \in N^+(d_i)} S_d(i,w)} \sum_{d_w \in N^+(d_i)} S_d(i,w) w_w, \quad (14)$$

where $N^+(d_i)$ denotes the set of neighbors of d_i composing of $k2$ known drugs, $S_d(i,w)$ denotes the similarity between d_i and a known drug d_w, w_w denotes the latent factor of d_w. Likewise, for the new disease p_j, the jth row of H is replaced in the same way. Thus, after inferring the latent factors for new drugs and diseases, the predicted association probability scores are calculated according to Equation (3).

2.4 CV on three benchmark datasets

In order to evaluate the performance of DRIMC in the three datasets we performed five trails of 10-fold CV under two settings. First, for each dataset, all known drug-disease associations were divided into 10 folds randomly with almost the same size. The 9-fold was regarded as positive training samples, while the remaining 1-fold was used as positive test samples. Notably, all the unobserved drug-disease associations were taken as negative test samples. Thus, the remaining 1-fold positive samples and all the unknown associations were deemed as test samples, and we aimed to identify the remaining 1-fold positive samples which should rank as the top N predictions. The process was repeated 10 times for each fold as a positive test set. Then, we conducted 10-fold CV for five times, each time with a different random seed. We denoted this CV setting as CVp. Second, we performed *de novo* drug-disease prediction to validate the capability of DRIMC in predicting potential indications for new drugs. Specifically, all drugs were randomly partitioned into 10 subsets with almost the same size. In each CV trail, the drug-disease associations with one subset were taken in turn as the test set, while the drug-disease associations with the remaining nine subsets constituted the training set. Likewise, we conducted 10-fold CV for five times, each time with a different random seed. We denoted this CV setting as CVd. Then, we evaluated the prediction performance by using the area under the ROC curve (AUC) and the area under the Precision-Recall (PR) curve (AUPR). We calculated an AUC (or AUPR) score in each of the 10 folds of every repetition and reported a final AUC (or AUPR) score that is the average over the five repeats of 10-fold CVs. It should be noted that there are not really being any negative samples, i.e. confirmed non-associations, so the ROC curve and PR curve are ranking the known associations on top of the unknowns rather than measuring exactly the prediction accuracy. However, because the true associations in reality are very rare compared with the total number of unknowns, the measurement properties of ROC and PR are still meaningful when the unknowns are regarded as true negatives.

3 Results

3.1 Similar therapeutic drugs have similar drug attributes

Here we defined the indication similarity between two drugs by using the maximum semantic similarity between their associated two disease groups. Furthermore, for each drug attribute type, we calculated the Jaccard coefficient between each pair of drug attribute profiles. For the PREDICT dataset, Figure 2 shows that chemical structure similarity, target domain similarity and target annotation similarity are significantly higher for drugs pairs having similar therapeutic indications. For drug pairs with indication similarity ≥ 0.5, the arithmetic mean values of their chemical structure similarity, target domain similarity and target annotation similarity are 0.212, 0.144 and 0.161, respectively. As a comparison, for drugs pairs with indication similarity <0.5, the arithmetic mean values for their above attribute similarity are 0.183, 0.093 and 0.113. Similar situations occur with the Cdataset and LRSSL dataset (see Supplementary Figure S1). These results suggest that drugs with similar indications tend to have similar attributes, such as chemical structures, target domains and target annotations.

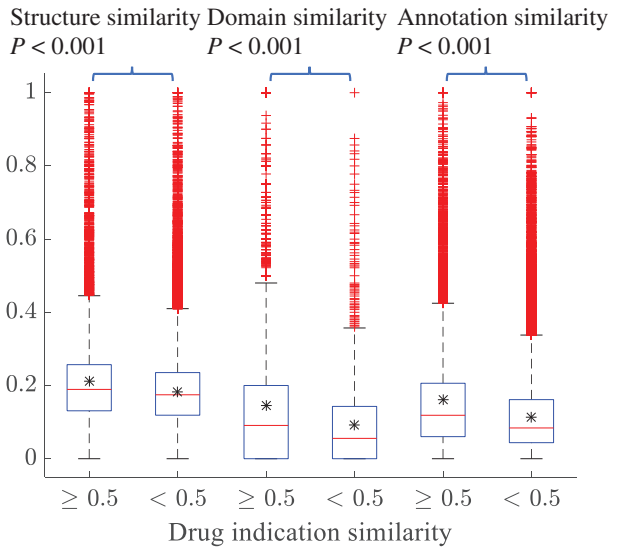

Fig. 2. Box plots of similar therapeutic drug pairs (indication similarity ≥ 0.5) and dissimilar therapeutic drug pairs (indication similarity < 0.5) with respect to their different attribute type similarity, including chemical structure similarity, target domain similarity and target annotation similarity, on PREDICT dataset. The asterisk in each box represents arithmetic mean

Furthermore, we found that some therapeutically similar drugs have dissimilar structures, while their target domains or target annotations are significantly similar. For the PREDICT dataset, Figure 3 shows the attribute similarity between pair of drugs having shared indications. Chemical structure similarity between every pair of drugs is <0.1, but either their target domain similarity or their target annotation similarity is >0.5. Nitroprusside and isosorbide dinitrate, which both can be used to treat renal failure, progressive, with hypertension, have a common target NPR1, and their target domain similarity is 1. Indapamide and guanethidine are also used in the treatment of renal failure, progressive, with hypertension, and they have similar target annotations (Jaccard coefficient 0.631). Pralidoxime is similar to pyridostigmine and neostigmine as they have shared targets ACHE and BCHE, and all these drugs are used in the treatment of myasthenia gravis. Exemestane is similar to letrozole and anastrozole because they share common target CYP19A1, all of which are used to treat breast cancer. These results show the necessity of combining drug target information for drug indication prediction.

3.2 Comparisons with the state-of-the-art algorithms

We compared DRIMC with four state-of-the-art methods for drug-disease association prediction, namely, DisDrugPred (Xuan *et al.*, 2019), SCMFDD (Zhang *et al.*, 2018), DRRS (Luo *et al.*, 2018) and MBiRW (Luo *et al.*, 2016). Furthermore, two matrix factorization-based methods for drug target prediction, i.e. kernelized Bayesian matrix factorization (KBMF, Gönen and Kaski, 2014) and neighborhood regularized logistic matrix factorization (NRLMF, Liu *et al.*, 2016), were also included in the comparison. As to drug repositioning prediction, KBMF can be directly fed with multiple views from the drug and disease data, while its prediction performance in this setting is significantly worse. Here we computed the average of the multiple similarity matrices for drugs (i.e. $\frac{1}{4}(S_d^1 + S_d^2 + S_d^3 + K_g^d)$) and the average of the two similarity matrices for diseases (i.e. $\frac{1}{2}(S_p^1 + K_g^p)$), and then fed them into the KBMF and NRLMF. The hyperparameters of DRIMC were set as following: the dimensionality of the subspace r was selected from $\{100, 150, 200, 250, 300\}$, regularization parameter λ was from $\{2^{-1}, 2^0, 2^1, 2^2\}$ and learning rate η was from $\{2^{-4}, 2^{-3}, 2^{-2}, 2^{-1}\}$, and then the set of parameters was obtained by using CV. When we set $r = 200$, $\lambda = 2$ and $\eta = 0.125$, we can get reasonably good results for all three benchmark datasets. DisDrugPred integrated four types of drug similarities according to drug features from multiple-view drugs, and one type of disease similarity based on their semantic associations. In DisDrugPred, regularization parameters $\alpha_1 \sim \alpha_6$ were selected from $\{0.05, 0.1, 0.2, 0.5, 1, 5, 10, 20, 50\}$ and then the optimal sets of parameters for different datasets were obtained by using grid search. SCMFDD, DRRS and MBiRW incorporated the drug chemical structure similarity, the disease semantic similarity and the known drug-disease associations for inferring novel drug indications. The hyperparameters of SCMFDD, DRRS and MBiRW were chosen as their optimal values provided by their publications. The hyperparameters of KBMF and NRLMF were determined by CV. In KBMF, the dimensionality of the subspace $R = 100$. For NRLMF, the dimensionality of the subspace $R = 200$, regularization parameters $\lambda_d = \lambda_t = 0.125$, $\alpha = 0.25$, $\beta = 0.125$ and learning rate $\gamma = 0.5$.

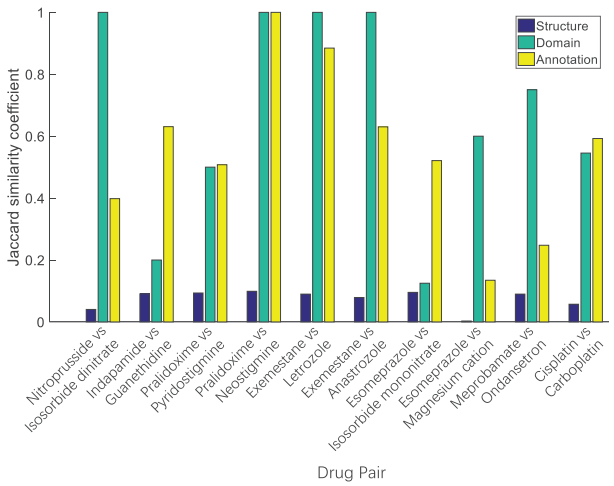

Fig. 3. Attribute similarity between pair of drugs having shared indications. The drug attributes contain chemical structure, target domain and target annotation

We first validated the prediction performance of DRIMC under the CV setting CVp. Table 1 shows the comparison results obtained by various methods. As shown in Table 1, DRIMC attains the best measure values in AUC over the three benchmark datasets. The AUC value obtained by DRIMC on PREDICT dataset is 0.956, which is 2.25% better than the second method NRLMF. On Cdataset, DRIMC obtains AUC 0.968, which is 2.11% better than the second method DRRS. On LRSSL dataset, DRIMC achieves AUC 0.954, which is 3.58% higher than the second method DisDrugPred. ROC curves of prediction results in one trail of 10-fold CV are illustrated in Supplementary Figure S2. For the AUPR metric, DRIMC achieves 0.299 and 0.377 for PREDICT dataset and Cdataset, which are 32.30% and 30.45% higher than the second method NRLMF, respectively. DRIMC achieves AUPR with 0.161 on LRSSL dataset, which is slightly lower than the best method NRLMF.

The comparison results obtained under the CV setting CVd for new drugs are shown in Table 2. DRIMC attains the best AUC values. The AUC values obtained by DRIMC over three datasets (i.e. PREDICT, Cdataset and LRSSL) are 0.873, 0.878 and 0.908, which are 3.31%, 3.29% and 4.13% better than the second competing method NRLMF, respectively. For the AUPR metric, DRIMC achieves the best result on LRSSL dataset. On PREDICT and Cdataset, NRLMF obtains a little better AUPR values than DRIMC.

To investigate the benefits of different drug attribute types contributing to prediction performance improvement, we exerted DRIMC on PREDICT dataset with different drug attribute combinations under CV settings CVp and CVd. Table 3 demonstrates the computational results. When only using drug chemical structure and GIP kernel, DRIMC achieves AUC 0.952 and AUPR 0.279 under the CV setting CVp. As to CV setting CVd, DRIMC obtains AUC 0.822 and AUPR 0.213 when using drug structure and GIP kernel only. The accuracy of DRIMC on Cdataset and LRSSL, obtained with combinations of drug attributes is shown in Supplementary Tables S1 and S2, respectively. Our results show that DRIMC

Table 1. The AUC and AUPR obtained under the CV setting CVp

Dataset	DisDrugPred	SCMFDD	MBiRW	DRRS	KBMF	NRLMF	DRIMC
AUC							
PREDICT	0.890	0.712	0.911	0.929	0.862	0.935	0.956
Cdataset	0.908	0.711	0.932	0.948	0.860	0.947	0.968
LRSSL	0.921	0.761	0.920	0.899	0.759	0.913	0.954
AUPR							
PREDICT	0.070	0.004	0.129	0.140	0.164	0.226	0.299
Cdataset	0.067	0.004	0.199	0.216	0.219	0.289	0.377
LRSSL	0.069	0.004	0.067	0.051	0.060	0.163	0.161

Table 2. The AUC and AUPR obtained under the CV setting CVd

Dataset	DisDrugPred	SCMFDD	MBiRW	DRRS	KBMF	NRLMF	DRIMC
AUC							
PREDICT	0.835	0.733	0.798	0.765	0.795	0.845	0.873
Cdataset	0.846	0.749	0.813	0.783	0.754	0.850	0.878
LRSSL	0.872	0.724	0.816	0.794	0.648	0.872	0.908
AUPR							
PREDICT	0.243	0.048	0.156	0.114	0.136	0.294	0.278
Cdataset	0.234	0.044	0.149	0.113	0.134	0.315	0.301
LRSSL	0.278	0.032	0.167	0.073	0.130	0.326	0.336

Table 3. Performance of DRIMC on PREDICT dataset obtained using combinations of drug attributes under the CV settings CVp and CVd

	CVp		CVd	
Attributes	AUC	AUPR	AUC	AUPR
Structure + GIP	0.952	0.279	0.822	0.213
Structure + GIP + Domain	0.955	0.291	0.856	0.243
Structure + GIP + Ontology	0.955	0.305	0.869	0.277
All	0.956	0.299	0.873	0.278

Structure, drug chemical structure; GIP, GIP kernel; Domain, drug target domain; Ontology, drug target annotation.

improves the prediction performance in both CV settings CVp and CVd thanks to combining target domain and target annotation, increasingly.

3.3 Advantage of neighborhood and sensitivity analysis

Our DRIMC method uses the k-nearest neighbors to select drug/disease features, and smooth latent factors of new drugs/diseases. As to the neighborhood sizes $k1$ and $k2$, they were determined by CV. When we set $k1 = 20$ and $k2 = 20$, we can get pretty good results for all three benchmark datasets. Here, we confirmed the neighborhood information benefits to prediction performance using PREDICT dataset under the setting CVp. When we set $k1 = 0$ (i.e. without consideration of neighborhood in drug/disease feature selection) and $k2 = 20$, the AUC and AUPR of DRIMC are 0.953 and 0.226, respectively. We varied $k1$ while keeping $k2 = 20$ fixed to verify that neighborhood-based feature selection can improve prediction accuracy. Figure 4A shows the performance trend of DRIMC, measured by AUC and AUPR with different settings of $k1$ under CVp. When $k1 = 20$, the AUPR is increased to 0.299 meanwhile the AUC is slightly better than that of $k1 = 0$. Likewise, by keeping $k1 = 20$ fixed, the accuracy trend of DRIMC including AUC and AUPR with respect to different settings of $k2$ is shown in Figure 4B. When we set $k1 = 20$ and $k2 = 0$ (i.e. without smoothness of the new drug/disease latent factors), the AUC and AUPR of DRIMC are 0.934 and 0.315, respectively. When $k2 = 20$, the AUC is increased to 0.956 meanwhile with the AUPR comparable. The trends of the accuracy of DRIMC on Cdataset and LRSSL, under the setting CVp with the different settings of neighborhood size are illustrated in Supplementary Figures S3 and S4, respectively. These results suggest that k-nearest neighbors could aid in drug/disease feature denoising and obtaining accurate association probability for a given drug-disease pair in prediction.

Next, we analyzed the impact of another hyperparameter, i.e. confidence of positive samples α, on the prediction accuracy. Figure 5 shows the prediction trend of DRIMC on PREDICT dataset, measured by AUC and AUPR under the setting CVp. The AUC and AUPR of DRIMC are 0.923 and 0.159 when $\alpha = 1$, while they are increased to 0.956 and 0.299 when $\alpha = 10$. The trends of the accuracy of DRIMC on Cdataset and LRSSL, under the setting CVp with the different settings of confidence are shown in Supplementary Figure S5. Finally, we performed sensitivity analysis on the dimensionality of the subspace r, and the initialization of the projection matrices W and H, and the results are shown in Supplementary Figures S6 and S7. We found that $r = 200$ obtains quite good results of AUC and AUPR measurements on three datasets, under the setting CVp. We initialized W and H using five different random seeds, and found that the results of prediction accuracy are almost the same.

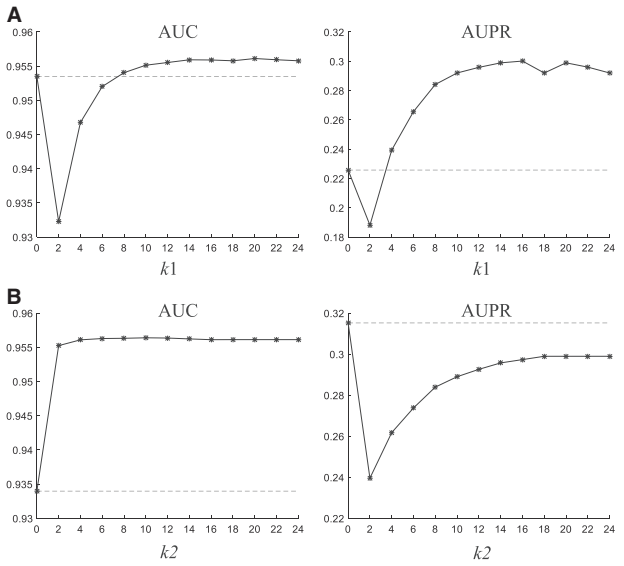

Fig. 4. Variation of the accuracy of DRIMC on PREDICT dataset, measured by AUC and AUPR under the CV setting CVp with the different settings of neighborhood size. (**A**) Varying $k1$ while keeping $k2 = 20$ fixed. (**B**)Varying $k2$ while keeping $k1 = 20$ fixed. $k1$ denotes the neighborhood size for selecting drug/disease features and $k2$ denotes the neighborhood size for smoothing new drug/disease latent factors

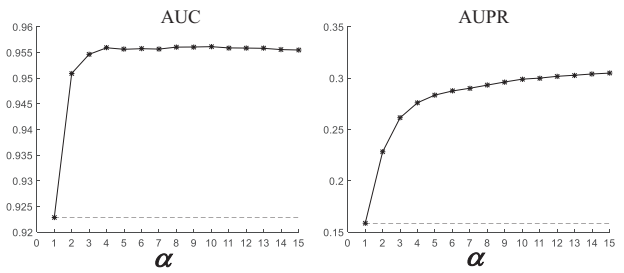

Fig. 5. Variation of the accuracy of DRIMC on PREDICT dataset, measured by AUC and AUPR under the CV setting CVp with the different settings of confidence α

3.4 Drug repositioning prediction

The newly predicted drug-disease associations can aid in drug repositioning. To reposition drugs to novel indications, all already

available drug-disease associations of LRSSL dataset were used as the training set. Then, the unknown associations will be ranked based on the prediction scores of DRIMC and the top ranked unknown associations were identified as the newly predicted associations. The comparative toxicogenomics database CTD (Davis *et al.*, 2019) was used as references to verify whether the newly predicted associations are true or not. CTD contains curated (marker/mechanism and therapeutic) and inferred drug-disease relationships, and only the therapeutic relationships were considered here for strict verification. Figure 6 shows the fractions of therapeutic effects among the top N ($N = 500$, 800, 1000, 1500) predictions generated by various drug repositioning methods, using the optimal parameters learned under CVp. We observed that the fractions of therapeutic effects achieved by DRIMC are 31.40%, 29.13%, 28.10% and 25.00% for top 500, 800, 1000 and 1500, respectively. The second method DRRS achieves 30.20%, 26.25%, 24.40% and 22.00% for top 500, 800, 1000 and 1500, respectively. Compared with other methods, DRIMC is able to achieve better prediction results across LRSSL dataset. Since the database is still being updated as new curated drug-disease associations are found, the fraction of new therapeutic effects correctly predicted by DRIMC may increase in the future. This satisfied result that DRIMC can successfully identify quite a few novel associations that are not in the LRSSL dataset, implies that DRIMC is capable of predicting new therapeutic effects from sparse matrices consisted of very few curated associations. This observation suggests that the proposed algorithm is very effective for finding novel therapeutic effects, thus it may help biologists or clinicians guide the experimental validations and reduce costs.

Furthermore, for each drug we analyzed the top N ($N = 5$, 10) predictions generated by our method DRIMC. We found that 635 drugs have at least one CTD curated therapeutic relationship which

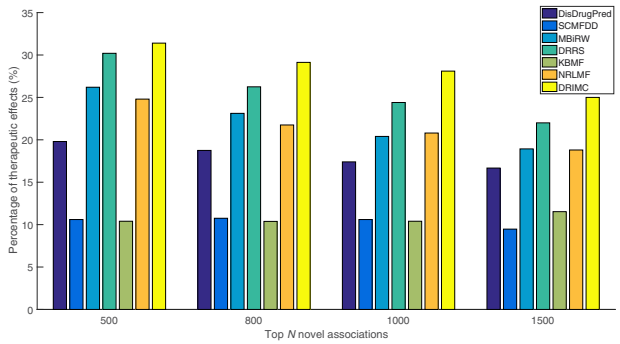

Fig. 6. Fractions of therapeutic effects among the top N predictions generated by seven methods for drug repositioning on LRSSL dataset. The CTD database is used as references to verify whether the newly predicted associations are therapeutic or not

does not appear in the training set (the set of relationships denoted by T). Among these drugs, 62.20% (395 out of 635) drugs have at least one therapeutic effect in T ranked in the top 5 by our method, 73.54% (467 out of 635) drugs have therapeutic effect in T ranked in the top 10. Some top-ranked predictions for drug repositioning are summarized in Table 4 and Supplementary Table S3. Besides several predicted therapeutic effects are confirmed by CTD records, some other predictions are validated by published literatures. Chloroquine is used to against giardiasis and osteoarthritis (Escobedo *et al.*, 2015; Rainsford *et al.*, 2015). Citalopram is used for the treatment of patients with bulimia nervosa (Leombruni *et al.*, 2006). Imatinib is an approved drug is used to treat certain types of cancer, but it is also found that this drug has potential for the treatment of Parkinson disease (Kumar, 2019). Ziprasidone is used for treating Huntington's disease in a clinical trial, and it is also helpful for controlling withdrawal signs in ethanol-dependent patients (Bonelli *et al.*, 2003; Celikyurt *et al.*, 2011). Warfarin has been the most prescribed anticoagulant in patients with atrial flutter for decades (Alalwan *et al.*, 2017). The antifungal drug fluconazole is indicated for leishmaniasis, paracoccidioidomycosis and lung diseases, fungal (Marques, 2012; Michelerio *et al.*, 2018; Skolnik *et al.*, 2017).

3.5 *De novo* drug indication prediction

We compiled an independent test set to perform *de novo* drug indication prediction. There are 61 new drugs which have at least one therapeutic indication curated in CTD database. The data sources for the new drugs consist of chemical structure, target domain and target ontology annotation which were obtained from Liang *et al.*'s (2017) work. Then, LRSSL dataset was used as training set and the *de novo* drug indication prediction was exerted by our method. For each new drug the top N ($N = 5$, 10) predictions were analyzed and validated by CTD database. We observed that 13.11% (8 out of 61) drugs have at least one therapeutic indication ranked in the top 5 by our method, 19.67% (12 out of 61) drugs have therapeutic indication ranked in the top 10.

Furthermore, for drugs and diseases in LRSSL dataset, we expanded the known drug-disease associations from 3051 to 8887 with the inclusion of the therapeutic records in CTD database, and which were used as training set again. We found that the percentage of drugs having at least one therapeutic effect ranked in the top 5, increased to 34.43% (21 out of 61), and the percentage of drugs having therapeutic effects ranked in the top 10, increased to 40.98% (25 out of 61). The top 5 predictions for some new drugs are summarized in Table 5 and Supplementary Table S4. Several predictions are confirmed by CTD records, while some other predictions are validated by published literatures. The inducible nitric oxide synthase (iNOS) is involved in the modulation of depressive behaviors, which could be abrogated remarkably by treatment with the iNOS inhibitor N-(3-(aminomethyl)benzyl)acetamidine (Peng *et al.*,

Table 4. Top 5 predictions for six randomly selected drugs based on drug repositioning

Drug	Disease
Chloroquine (DB00608)	Giardiasis (D005873); Malaria (D008288)[a]; Malaria, Vivax (D016780)[a]; Osteoarthritis (D010003); Peripheral Vascular Diseases (D016491)
Citalopram (DB00215)	Anxiety Disorders (D001008)[a]; Attention Deficit Disorder with Hyperactivity (D001289)[a]; Autistic Disorder (D001321)[a]; Bulimia (D002032); Panic Disorder (D016584)[a]
Imatinib (DB00619)	Acquired Immunodeficiency Syndrome (D000163); Breast Neoplasms (D001943)[a]; Parkinson Disease (D010300); Leukemia, Myelogenous, Chronic, BCR-ABL Positive (D015464)[a]; Stomach Neoplasms (D013274)[a]
Ziprasidone (DB00246)	Autistic Disorder (D001321)[a]; Bipolar Disorder (D001714)[a]; Huntington Disease (D006816); Psychotic Disorders (D011618)[a]; Substance Withdrawal Syndrome (D013375)
Warfarin (DB00682)	Atrial Flutter (D001282); Embolism (D004617)[a]; Myocardial Infarction (D009203)[a]; Thrombosis (D013927)[a]; Venous Thrombosis (D020246)[a]
Fluconazole (DB00196)	Aspergillosis (D001228)[a]; Candidiasis, Cutaneous (D002179)[a]; Leishmaniasis (D007896); Lung Diseases, Fungal (D008172); Paracoccidioidomycosis (D010229)

[a]Therapeutic effect which is verified by CTD.

Table 5. Top 5 predictions for six new drugs

Drug	Disease
Bivalirudin (DB00006)	Postoperative Complications (D011183); Thrombocytopenia (D013921)[a]; Thromboembolism (D013923); Thrombosis (D013927)[a]; Venous Thrombosis (D020246)
N-(3-(aminomethyl)benzyl) acetamidine (DB02044)	Depressive Disorder (D003866); Edema (D004487)[a]; Hypertension (D006973); Pain (D010146); Substance Withdrawal Syndrome (D013375)
Quercetin (DB04216)	Anxiety Disorders (D001008)[a]; Edema (D004487)[a]; Inflammation (D007249)[a]; Pain (D010146)[a]; Seizures (D012640)
Nomifensine (DB04821)	Anxiety Disorders (D001008); Bipolar Disorder (D001714); Depressive Disorder (D003866)[a]; Pain (D010146); Panic Disorder (D016584)
Farnesol (DB02509)	Anxiety Disorders (D001008); Bipolar Disorder (D001714); Depressive Disorder (D003866); Parkinson Disease (D010300); Tremor (D014202)
Fisetin (DB07795)	Depressive Disorder (D003866); Hypertension (D006973); Inflammation (D007249); Pain (D010146); Seizures (D012640)

[a]Therapeutic effect which is verified by CTD.

2012). Besides, N-(3-(aminomethyl)benzyl)acetamidine can limit pain hypersensitivity in a neuropathic pain rat model (Staunton *et al.*, 2018). Quercetin has shown strong anti-epileptic effect in animal models (Singh et al., 2017). Farnesol has potential treatment of anxiety disorder and Parkinson's disease (Sari and Khalil, 2015; Shahnouri *et al.*, 2016). Fisetin, a polyphenolic compound, has drawn notable attention owing to its anti-depressant, anti-inflammatory, anti-epileptic and cardioprotective effects (Pal *et al.*, 2016).

4 Conclusion

In this article, we developed a computational drug repositioning method based on DRIMC. DRIMC has the following advantages over other state-of-the-art drug repositioning approaches. First, heterogeneous drug and disease data sources are integrated into side information which can be fed into the model directly. Second, the model assigns higher confidence levels to known association pairs compared with unknown pairs so as to take advantage of the sparsity of the drug-disease association network. Third, the model focuses on the local structure of the drug-disease association data, by utilizing the neighborhood effects from most similar drugs and most similar diseases. Specifically, the model uses nearest neighbors to create drug and disease side information, and smooth latent factors of new drugs and new diseases. In this way our method only exploits nearest neighbors instead of all similar neighbors as considered in previous studies, and thus gets more accurate results by avoiding noisy information.

The performance of DRIMC was validated by the five trails of 10-fold CV on three benchmark datasets. Clearly, DRIMC shows better overall prediction accuracy than other methods in the comparison study. Finally, using the already available drug-disease association data, both the drug repositioning prediction and the *de novo* drug prediction can find novel drug-disease associations which are verified by the publicly available database CTD or literatures. DRIMC can serve as a tool to identify candidate indications for existing drugs which can be used to guide following wet experiments. Besides, DRIMC can be applied to other research fields such as synergistic drug combination prediction (Chen *et al.*, 2016). Drug combination therapy can be regarded as a new route for repositioning old drugs. By incorporating features associated with two drugs in matrix completion, DRIMC enables prediction for drug combination.

Funding

This work was supported by National Natural Science Foundation of China [61603273]; Tianjin Municipal Natural Science Foundation [18JCQNJC69500]; Scientific Research Program of Tianjin Education Commission [2018KJ107]; and MOE (Ministry of Education in China) Youth Foundation of Humanities and Social Sciences [18YJC630108].

Conflict of Interest: none declared.

References

Alalwan,A.A. *et al.* (2017) Trends in utilization of warfarin and direct oral anticoagulants in older adult patients with atrial fibrillation. *Am. J. Health Syst. Pharm.*, **74**, 1237–1244.

Bonelli,R.M. *et al.* (2003) Ziprasidone in Huntington's disease: the first case reports. *J. Psychopharmacol.*, **17**, 459–460.

Celikyurt,I.K. *et al.* (2011) Effects of risperidone, quetiapine and ziprasidone on ethanol withdrawal syndrome in rats. *Prog. Neuropsychopharmacol. Biol. Psychiatry*, **35**, 528–536.

Chen,X. *et al.* (2016) NLLSS: predicting synergistic drug combinations based on semi-supervised learning. *PLoS Comput. Biol.*, **12**, e1004975.

Chen,X. *et al.* (2018) Predicting miRNA-disease association based on inductive matrix completion. *Bioinformatics*, **34**, 4256–4265.

David,S.W. *et al.* (2018) DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.*, **46**, D1074–D1082.

Davis,A.P. *et al.* (2019) The comparative toxicogenomics database: update 2019. *Nucleic Acids Res.*, **47**, D948–D954.

Duchi,J. *et al.* (2011) Adaptive subgradient methods for online learning and stochastic optimization. *J. Mach. Learn. Res.*, **12**, 2121–2159.

Escobedo,A.A. *et al.* (2015) Chloroquine: an old drug with new perspective against giardiasis. *Recent Pat. Antiinfect. Drug Discov.*, **10**, 134–141.

Gönen,M., and Kaski,S. (2014) Kernelized Bayesian matrix factorization. *IEEE Trans. Pattern Anal. Mach. Intell.*, **36**, 2047–2060.

Gottlieb,A. *et al.* (2011) PREDICT: a method for inferring novel drug indications with application to personalized medicine. *Mol. Syst. Biol.*, **7**, 496.

Hamosh,A. *et al.* (2002) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.*, **30**, 52–55.

Hao,M. *et al.* (2016) Improved prediction of drug-target interactions using regularized least squares integrating with kernel fusion technique. *Anal. Chim. Acta*, **909**, 41–50.

Hao,M. *et al.* (2017) Predicting drug-target interactions by dual-network integrated logistic matrix factorization. *Sci. Rep.*, **7**, 40376.

Jain,P. and Dhillon,I.S. (2013) Provable inductive matrix completion. *CoRR*, abs/1306.0626.

Johnson,C.C. (2014) Logistic matrix factorization for implicit feedback data. *Advances in Neural Information Processing Systems, 27*.

Kumar,M. *et al.* (2019) Expanding spectrum of anticancer drug, imatinib, in the disorders affecting brain and spinal cord. *Pharmacol. Res.*, **143**, 86–96.

Leombruni,P. *et al.* (2006) Citalopram versus fluoxetine for the treatment of patients with bulimia nervosa: a single-blind randomized controlled trial. *Adv. Ther.*, **23**, 481–494.

Liang,X. *et al.* (2017) LRSSL: predict and interpret drug-disease associations based on data integration using sparse subspace learning. *Bioinformatics*, **33**, 1187–1196.

Liu,Y. *et al.* (2016) Neighborhood regularized logistic matrix factorization for drug-target interaction prediction. *PLoS Comput. Biol.*, **12**, e1004760.

Luo,H. *et al.* (2016) Drug repositioning based on comprehensive similarity measures and bi-random walk algorithm. *Bioinformatics*, **32**, 2664–2671.

Luo,H. *et al.* (2018) Computational drug repositioning using low-rank matrix approximation and randomized algorithms. *Bioinformatics*, **34**, 1904–1912.

Marques,S.A. (2012) Paracoccidioidomycosis. *Clin. Dermatol.*, **30**, 610–615.

Martínez,V. *et al*. (2015) DrugNet: network-based drug-disease prioritization by integrating heterogeneous data. *Artif. Intell. Med*., **63**, 41–49.

Michelerio,A. *et al*. (2018) Pediatric old world cutaneous leishmaniasis treated with oral fluconazole: a case series. *Pediatr. Dermatol*., **35**, 384–387.

Natarajan,N. and Dhillon,I. (2014) Inductive matrix completion for predicting gene-disease associations. *Bioinformatics*, **30**, i60–i68.

Oh,M. *et al*. (2014) A network-based classification model for deriving novel drug-disease associations and assessing their molecular actions. *PLoS One*, **9**, e111668.

Pal,H.C. *et al*. (2016) Fisetin and its role in chronic diseases. *Adv. Exp. Med. Biol*., **928**, 213–244.

Paul,S.M. *et al*. (2010) How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat. Rev. Drug Discov*., **9**, 203–214.

Peng,Y.L. *et al*. (2012) Inducible nitric oxide synthase is involved in the modulation of depressive behaviors induced by unpredictable chronic mild stress. *J. Neuroinflammation*, **9**, 75.

Persidis,A. *et al*. (2011) The benefits of drug repositioning. *Drug Discov*., **12**, 9–12.

Pushpakom,S. *et al*. (2019) Drug repurposing: progress, challenges and recommendations. *Nat. Rev. Drug Discov*., **18**, 41–58.

Rainsford,K.D. *et al*. (2015) Therapy and pharmacological properties of hydroxychloroquine and chloroquine in treatment of systemic lupus erythematosus, rheumatoid arthritis and related diseases. *Inflammopharmacology*, **23**, 231–269.

Sari,Y. and Khalil,A. (2015) Monoamine oxidase inhibitors extracted from tobacco smoke as neuroprotective factors for potential treatment of Parkinson's disease. *CNS Neurol. Disord. Drug Targets*, **14**, 777–785.

Shahnouri,M. *et al*. (2016) Neuropharmacological properties of farnesol in Murine model. *Iran J. Vet. Res*., **17**, 259–264.

Singh,T. *et al*. (2017) Adjuvant quercetin therapy for combined treatment of epilepsy and comorbid depression. *Neurochem. Int*., **104**, 27–33.

Skolnik,K. *et al*. (2017) Cryptococcal lung infections. *Clin. Chest Med*., **38**, 451–464.

Staunton,C.A. *et al*. (2018) Inducible nitric oxide synthase inhibition by 1400W limits pain hypersensitivity in a neuropathic pain rat model. *Exp. Physiol*., **103**, 535–544.

Van Driel,M.A. *et al*. (2006) A text-mining analysis of the human phenome. *Eur. J. Human Genet*., **14**, 535–542.

Wang,B. *et al*. (2014a) Similarity network fusion for aggregating data types on a genomic scale. *Nat. Methods*, **11**, 333–337.

Wang,L. *et al*. (2017) Improved anticancer drug response prediction in cell lines using matrix factorization with similarity regularization. *BMC Cancer*, **17**, 513.

Wang,W. *et al*. (2014b) Drug repositioning by integrating target information through a heterogeneous network model. *Bioinformatics*, **30**, 2923–2930.

Wang,Y. *et al*. (2013) Drug repositioning by kernel-based integration of molecular structure, molecular activity, and phenotype data. *PLoS One*, **8**, e78518.

Xuan,P. *et al*. (2019) Drug repositioning through integration of prior knowledge and projections of drugs and diseases. *Bioinformatics*, **35**, 4108–4119.

Yang,L. and Agarwal,P. (2011) Systematic drug repositioning based on clinical side-effects. *PLoS One*, **6**, e28025.

Yap,C.W. (2011) PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. *J. Comput. Chem*., **32**, 1466–1474.

Zhang,W. *et al*. (2018) Predicting drug-disease associations by using similarity constrained matrix factorization. *BMC Bioinformatics*, **19**, 233.