

RESEARCH

Open Access



In silico drug repositioning based on integrated drug targets and canonical correlation analysis

Hailin Chen^{1*}, Zuping Zhang² and Jingpu Zhang³

Abstract

Background: Besides binding to proteins, the most recent advances in pharmacogenomics indicate drugs can regulate the expression of non-coding RNAs (ncRNAs). The polypharmacological feature in drugs enables us to find new uses for existing drugs (namely drug repositioning). However, current computational methods for drug repositioning mainly consider proteins as drug targets. Meanwhile, these methods identify only statistical relationships between drugs and diseases. They provide little information about how drug-disease associations are formed at the molecular target level.

Methods: Herein, we first comprehensively collect proteins and two categories of ncRNAs as drug targets from public databases to construct drug–target interactions. Experimentally confirmed drug-disease associations are downloaded from an established database. A canonical correlation analysis (CCA) based method is then applied to the two datasets to extract correlated sets of targets and diseases. The correlated sets are regarded as canonical components, and they are used to investigate drug's mechanism of actions. We finally develop a strategy to predict novel drug-disease associations for drug repositioning by combining all the extracted correlated sets.

Results: We receive 400 canonical components which correlate targets with diseases in our study. We select 4 components for analysis and find some top-ranking diseases in an extracted set might be treated by drugs interfacing with the top-ranking targets in the same set. Experimental results from 10-fold cross-validations show integrating different categories of target information results in better prediction performance than only using proteins or ncRNAs as targets. When compared with 3 state-of-the-art approaches, our method receives the highest AUC value 0.8576. We use our method to predict new indications for 789 drugs and confirm 24 predictions in the top 1 predictions.

Conclusions: To the best of our knowledge, this is the first computational effort which combines both proteins and ncRNAs as drug targets for drug repositioning. Our study provides a biologically relevant interpretation regarding the forming of drug-disease associations, which is useful for guiding future biomedical tests.

Keywords: Drug repositioning, Canonical correlation analysis, Integrated targets

Background

Over 100 years ago, the Nobel laureate Paul Ehrlich established his revolutionary ‘magic bullet’ concept, which has successfully inspired generations of chemists and pharmacologists to create target-specific drugs for disease treatment [1]. This declared paradigm has become a pragmatic criterion in drug discovery for the

*Correspondence: chen hailin@ecjtu.edu.cn

¹ School of Software, East China Jiaotong University, Nanchang 330013, China

Full list of author information is available at the end of the article



past decades. However, the interpretation of the magic bullet as a drug which acts through a single crucial target in an exclusive and highly specific way has been challenged, because increasing studies demonstrate drugs usually have multiple physiological targets rather than one target [2–4].

The polypharmacological feature in drugs enables us to find new indications (also known as drug repositioning [5]) for existing drugs. For instance, a study conducted by Skrott et al. [6] found that the metabolite of disulfiram binds to a new target NPL4, which is responsible for anti-cancer effects. Therefore, the old alcohol-aversion drug can be repurposed for tumour treatment. Meanwhile, unintended ‘off-targets’ may cause adverse drug reactions (ADR) [7], which would limit the use of drugs. It is therefore necessary to discover the real targets implicated in drug indications.

There are 4 potential types of macromolecules in biological systems with which we can interfere using small-molecule drugs: proteins, polysaccharides, lipids and nucleic acids [8]. Previous research efforts were mainly made on the first type of molecular targets [9–12]. The most recent studies in pharmacogenomics have discovered that drugs can regulate the expression levels of two categories of ncRNAs, namely miRNAs and lncRNAs. For example, Smith et al. [13] revealed that the expression levels of 44 miRNAs are repressed during glucocorticoid-induced apoptosis. Guo et al. [14] identified aspirin can activate the expression of a lncRNA named *OLAIP2* in human colorectal cancer. Given the intriguing fact that ncRNAs play significant roles in disease development [15–17], targeting these ncRNAs with small-molecule drugs offers another new and promising type of therapy for human diseases [18–23].

As traditional biomedical experiments are expensive and time-consuming, computational approaches provide an alternative tool for drug repositioning. For example, Chen et al. [24] exploited multiple heterogeneous data to integrate drug-disease network and drug–target network into one coherent model, and applied cross-network embedding to predict drug-disease associations for drug repositioning. A comprehensive and detailed survey on computational drug repositioning is available at Review [25]. Note that previous computational approaches for drug repositioning seldom take integrated target information into consideration. They usually exploit proteins as drug targets. We argue that integrating different types of targets would provide a better and more comprehensive understanding of drug’s MoA. Further, these methods discover only statistical associations between drugs and diseases at data level. They seldom investigate how drug-disease associations are formed at the molecular target level.

Table 1 Statistics of the drug–target interactions used in our manuscript

Name	Statistics
# drugs	1190
# total targets (including proteins, miRNAs and lncRNAs)	1668
# proteins	1167
# miRNAs	348
# lncRNAs	153
# total drug–target interactions	5331
# drug–protein interactions	4337
# drug–miRNA interactions	825
# drug–lncRNA interactions	169
Average number of targets for each drug	4.5

Table 2 Statistics of the drug-disease associations used in our manuscript

Name	Statistics
# drugs	1190
# diseases	1111
# drug-disease associations	5869
Average number of associated diseases for each drug	4.9

In this paper, we first comprehensively select drug targets from proteins, miRNAs and lncRNAs to construct drug–target interactions. Therapeutically verified drug indications are downloaded to form drug-disease associations. Then, we apply a CCA-based method to extract correlated sets of targets and diseases. The correlated targets and diseases provide explanations of the forming of drug-disease associations. We finally predict novel drug-disease associations for drug repositioning by combining the correlated sets. Comprehensive experiments demonstrate using integrated target information not only improves prediction performance, but also provides a more extensive view of drug’s MoA. Case studies suggest some top predictions are confirmed by existing databases. When compared with other methods using the benchmark datasets in our study, our approach shows improvements in terms of AUC value.

Results

Preliminary analysis of the datasets

In total, we receive 1190 drugs with both target and indication information. For the 1190 drugs, we obtain 5331 drug–target interactions containing 1668 targets and 5869 drug-disease associations including 1111 diseases. An overview of the two datasets is available at Tables 1 and 2, respectively.

We further use a boxplot (Additional file 1) to describe the distribution of numbers of targets and indications of the 1190 drugs. We discover that there are 885 (74.4%) drugs whose target numbers are less than 4.5 (the average value) and 887 (80.0%) drugs whose indication numbers are less than 4.9 (the average value). Meanwhile, as a category of newly discovered targets, the number of experimentally supported drug–ncRNA interactions are far less than that of drug–protein interactions. We can conclude from the analysis that our knowledge about drug–target interactions and drug-disease associations is not complete.

Performance evaluation

In this study, we collect both proteins and ncRNAs as drug targets. We therefore separately use proteins, ncRNAs and integrated targets to conduct 10-fold cross-validation experiments. We use average AUC values for performance evaluation. The results are summarized in Table 3. We discover that integrating both proteins and ncRNAs results in better prediction performance than only using proteins or ncRNAs as targets. We also find that imposing sparsity constraint on CCA can improve prediction performance. Note almost all elements in the weight vectors in ordinary CCA (OCCA) are non-zero, indicating that OCCA cannot select a small number of features as informative drug targets and indications.

Effects of parameters on cross-validation experiments

There are three parameters (c_1 , c_2 and k) in our method. The parameters c_1 and c_2 are to control the sparsity. The parameter k is the number of canonical components. For simplicity, we choose the same value for c_1 and c_2 . We comprehensively set the values of c_1 and c_2 in the range of

[0.1, 0.9], and the value of k in the range of [60, 500] when conducting 10-fold cross validations. We list the average AUC values in Table 4. We find the best inference performance is achieved when $c_1 = c_2 = 0.1$, and $k = 400$.

Investigating drug’s MoA at the molecular target level

Drugs exert their therapeutic effects through modulating their biological targets, and in turn promote healthy functioning of our metabolic system. As a drug usually has multiple targets, detecting the real target(s) implicated in a disease is critical for understanding drug’s MoA and for further drug repositioning.

We obtain 400 canonical components (Additional file 2) which correlate targets with diseases. We use four components (#1, #3, #6 and #7) as examples to investigate the biological meaning of the extracted sets of targets and diseases. We select the top targets and diseases in each component for analysis.

In component #1, there are 34 targets and 23 diseases with positive weight. We find from the database DisGeNET [26] that 4 high-ranking target proteins, *Interleukin-1 beta* (3rd), *Caspase-1* (3rd), *Caspase-3* (3rd) and *Matrix metalloproteinase-9* (3rd), are associated with the top disease *Periodontitis* (1st). Two top-scoring targets, *Interleukin-1 beta* (3rd) and *Matrix metalloproteinase-9* (3rd), are related with one top-scoring disease *Cholera* (4th). The target *Caspase-3* (3rd) is associated with the disease *Chlamydia trachomatis infection of genital structure* (5th).

Similar findings are discovered in component #3, #6 and #7. We list the confirmed top target-disease associations in the three components in Additional file 3, 4 and 5, respectively. Besides proteins, ncRNAs are found to be associated with diseases. For example, we discover

Table 3 Average AUC values received from the CCA methods based on 10-fold cross-validations

	SCCA (proteins + ncRNAs)	SCCA (ncRNAs)	SCCA (proteins)	OCCA (proteins + ncRNAs)	OCCA (ncRNAs)	OCCA (proteins)
AUC value	0.8576	0.7391	0.8537	0.8107	0.7283	0.8106

The bold value indicated the highest one

Table 4 Average AUC values received based on 10-fold cross-validations by parameter tuning

	$k = 60$	80	100	200	300	400	500
$c_1 = c_2 = 0.1$	0.8146	0.8244	0.8293	0.8463	0.8542	0.8576	0.8575
0.3	0.8124	0.8124	0.8107	0.8027	0.8012	0.8014	0.8003
0.5	0.8146	0.8099	0.8026	0.7753	0.7717	0.7686	0.7649
0.7	0.8160	0.8107	0.8043	0.7752	0.7702	0.7659	0.7645
0.9	0.8160	0.8106	0.8042	0.7751	0.7702	0.7659	0.7645

The bold value indicated the highest one

in component #3 the top-ranking miRNA (*miR-135b*) is related with *malignant neoplasm of thyroid* (4th), *malignant neoplasm of lung* (6th) and *breast carcinoma* (7th), and the top-ranking miRNA (*miR-520h*) is associated with *malignant neoplasm of lung* (6th) and *breast carcinoma* (7th). These relationships are confirmed by the database HMDD [15]. In component #7, a lncRNA *UCA1* (8th) is found to be related with *Leukemia, Myeloid, Chronic-Phase* (3rd), which is verified by the database LncRNADisease [16]. Based on these findings, we presume drugs may act on the top-ranking targets in one canonical component to treat the top-ranking diseases in the same component.

Comparison with other methods

As mentioned before, this is the first computational effort using integrated targets for drug repositioning. Previous computational approaches for drug repositioning were developed based on different data features they analysed. We therefore choose 3 other methods which can take our datasets as inputs for comparison. The 3 baseline methods are as follows:

- DBSI [27]: a collaborative-filtering-based method using chemical similarity for drug–target interaction prediction.
- SDTNBI [28]: an integrated tool for large-scale drug–target interaction prediction using chemical substructures.
- MLKNN [29]: a multi-label k-nearest neighbour method for drug side effect prediction.

To make fair comparison, we apply the 3 methods to our datasets and use 10-fold cross-validations for prediction performance comparison. For the method DBSI, we calculate drug–drug similarity according to Jaccard score based on their target information. This strategy of similarity calculation has been applied in other studies [30, 31]. The received AUC values for these methods

are shown in Table 5. We perform Wilcoxon rank sum tests between SCCA and the other 3 methods based on the AUC values. The calculated *p* values are available at Table 6. The experimental results demonstrate our approach SCCA performs best in the 4 methods. Note that the other 3 methods cannot provide clues for biological interpretation.

New indication prediction for existing drugs

After confirming the prediction ability, we further apply our method to those drugs, which are not in the benchmark datasets but whose target information is available, for their new indication predictions. There are 789 drugs of such kind. All known information, including drug–target interactions and drug–disease associations, in our gold-standard datasets is used for training. The potential indications are prioritized based on the prediction scores in descending order according to the method SCCA.

We list the top 50 predicted results of the 789 drugs in Additional file 6 for future screenings. We further validate the top *k* (*k*=5, 10, 20, 30 and 50) predictions by checking the public database CTD [32], a knowledgebase that houses information of chemicals, genes, phenotypes, diseases and exposures to advance understanding about human health. As this database contains both inferred and curated records, we only select curated drug–disease associations for prediction confirmation. The numbers of confirmed drug indications in the top *k* predictions are illustrated in Fig. 1. Because of space limitation, we only report the top 1 drug indication predictions supported by CTD in Table 7. More detailed information of the verified drug–disease associations in the top 50 predictions is available at Additional file 7. The excellent results indicate our method can be applied in real situations.

Discussion

Uncovering drug's MoA is of great importance for drug repositioning. In vivo and in vitro experiments are useful but expensive tools to address the problem. Our

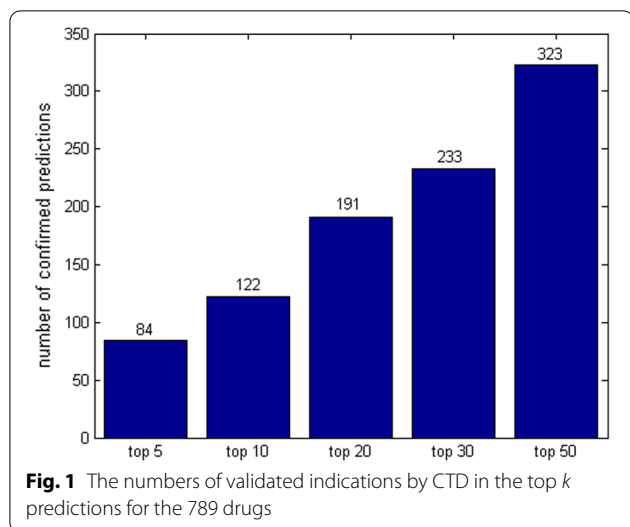
Table 5 Comparison of average AUC values with existing methods based on 10-fold cross-validations

	SCCA	DBSI	SDTNBI	MLKNN
AUC value	0.8576 ± 0.0005	0.8413 ± 0.0022	0.8395 ± 0.0010	0.7945 ± 0.0002

The bold value indicated the highest one

Table 6 The *p*-values received from Wilcoxon rank sum tests

	DBSI	SDTNBI	MLKNN
<i>p</i> value between SCCA and another method based on AUC values	1.6305E−04	1.7168E−04	1.6973E−04



CCA-based computational method provides an alternative to revealing the targets which are implicated in drug indications, and results suggest the extracted sets of targets and diseases are biologically meaningful. Compared with previous studies, we integrate both proteins

and ncRNAs as drug targets. Experiments further demonstrate using integrated targets improves prediction performance.

Even though, our proposed method has been shown to be useful in drug repositioning. Some limitations in this study need to be pointed out. First, our method depends heavily on known drug–target interactions and drug–disease associations. As we know, many drug targets (especially drug–ncRNA interactions) and drug indications have not been discovered. The incompleteness of data would result in biased prediction results. We expect combining more experimentally confirmed drug–target interactions and drug–disease associations would provide more reliable predictions. Meanwhile, there are 3 parameters in our method. Selecting appropriate values for the 3 parameters to receive optimal results is a challenging task. Third, the numbers of extracted components are determined by the parameter k (see Eq. 4) in our method, and different numbers of extracted components would influence our interpretation of drug’s MoA.

More recently, a growing number of studies [24, 33–43] are exploiting both features from drugs and diseases for drug repositioning. Integrating these features may provide more reliable prediction results. Another trend

Table 7 The confirmed results in the top 1 drug indication predictions by CTD

Drug name	Disease name	Ranking in the prediction list	Evidence
Troglitazone	Hypertriglyceridemia	Top 1	CTD
Methysergide	Migraine disorders	Top 1	CTD
Ropivacaine	Pruritus	Top 1	CTD
Tenofovir disoproxil	HIV infections	Top 1	CTD
Remoxipride	Schizophrenia	Top 1	CTD
Rosiglitazone	Hypercholesterolemia	Top 1	CTD
Cerivastatin	Hypercholesterolemia	Top 1	CTD
Meperidine	Pain	Top 1	CTD
Dronabinol	Obesity	Top 1	CTD
Phenindione	Thromboembolism	Top 1	CTD
Amodiaquine	Malaria, falciparum	Top 1	CTD
Alfentanil	Pain	Top 1	CTD
Risedronic acid	Osteoporosis, postmenopausal	Top 1	CTD
Levobupivacaine	Pruritus	Top 1	CTD
Ketamine	Pain	Top 1	CTD
Sulfadoxine	Malaria, falciparum	Top 1	CTD
Methotrimeprazine	Schizophrenia	Top 1	CTD
Acenocoumarol	Thromboembolism	Top 1	CTD
Diamorphine	Pain	Top 1	CTD
Pimavanserin	Schizophrenia	Top 1	CTD
Ciprofibrate	Hypertriglyceridemia	Top 1	CTD
Vitamin d	Hypoparathyroidism	Top 1	CTD
Elagolix	Endometriosis	Top 1	CTD
mg132	Multiple myeloma	Top 1	CTD

in drug repositioning is drug combinations [44–46] (see Review [47] for more details), which can result in low adverse side effects and high treatment efficacy compared to single drug administration. We believe these efforts offer help with drug discovery and disease treatment from different perspectives.

Conclusions

In this study, we apply a CCA-based method to extract correlated sets of targets and diseases, and the correlated targets and diseases provide clues for explaining drug’s MoA for drug repositioning. We further propose a prediction scheme for drug repositioning based on the extracted correlated sets. Experimental results of cross-validations indicate that integrating different categories of targets and imposing sparsity constraint on CCA improve prediction performance. Case studies demonstrate that some of the top predictions by our method are supported by literature. Moreover, our method shows improvement in prediction accuracy when compared with other approaches. We expect that our study offers a useful tool for drug repositioning.

Methods

Data preparation

We collect two datasets, namely drug–target interactions and drug-disease associations, from public databases for our study. The two datasets are regarded as gold-standard data. We use the benchmark datasets to evaluate the performance of our method. We also use the two datasets as training datasets for comprehensive indication prediction.

For drug–target interactions, we integrate 3 categories of macromolecules (proteins, miRNAs and lncRNAs) as drug targets. We obtain drug–protein interactions from DrugBank [48], a freely available web resource containing detailed information about drugs, their mechanisms, their interactions and their targets. We only select small molecule drugs and approved targets in DrugBank in our study. We download drug–miRNA interactions and drug–lncRNA interactions from SM2miR [49] and D-lnc [50], respectively. The two databases separately provide comprehensive repositories to detect the modification of drugs on miRNA and lncRNA expression. We restrict the species to *Homo sapiens* in both databases. We do not take inferred results in D-lnc for consideration.

Drug-disease associations are received from repoDB [51], a database consisting of approved and failed drugs and their indications. We only keep the approved drug-indication pairs in the database in our datasets.

Method description

Suppose that we have a set of m drugs with p molecular target features and q disease features. We denote each drug by a target feature vector $t = (t_1, t_2, t_3, \dots, t_p)^T$ and by a disease feature vector $d = (d_1, d_2, d_3, \dots, d_q)^T$, where t_i (or d_j) is represented for the presence or absence of a target (or a disease) by 1 or 0, respectively.

Consider two linear combinations for targets and diseases as $u_i = \alpha^T t_i$ and $v_i = \beta^T d_i (i = 1, 2, 3, \dots, m)$, where $\alpha = (\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_p)^T$ and $\beta = (\beta_1, \beta_2, \beta_3, \dots, \beta_q)^T$ are weight vectors. We apply canonical correlation analysis [52] to find weight vectors α and β which maximize the following correlation coefficient:

$$\rho = \text{corr}(u, v) = \frac{\sum_{i=1}^m \alpha^T t_i \cdot \beta^T d_i}{\sqrt{\sum_{i=1}^m (\alpha^T t_i)^2} \sqrt{\sum_{i=1}^m (\beta^T d_i)^2}} \tag{1}$$

Let X denote an $m \times p$ matrix and Y denote an $m \times q$ matrix. Then the maximization problem can be formally rewritten as follows:

$$\max \text{imize} \{ \alpha^T X^T Y \beta \} \text{ subject to } \|\alpha\|_2^2 \leq 1 \|\beta\|_2^2 \leq 1. \tag{2}$$

We refer to it as ordinary canonical correlation analysis (OCCA). OCCA usually results in vectors α and β with many non-zero elements. To impose sparsity on α and β , we choose to add penalties to (2) like reference [53–55] and the maximization problem is considered as:

$$\begin{aligned} \max \text{imize} \{ \alpha^T X^T Y \beta \} \\ \text{subject to } \|\alpha\|_2^2 \leq 1, \|\beta\|_2^2 \leq 1, \|\alpha\|_1 \leq c_1 \sqrt{p} \|\beta\|_1 \leq c_2 \sqrt{q} \end{aligned} \tag{3}$$

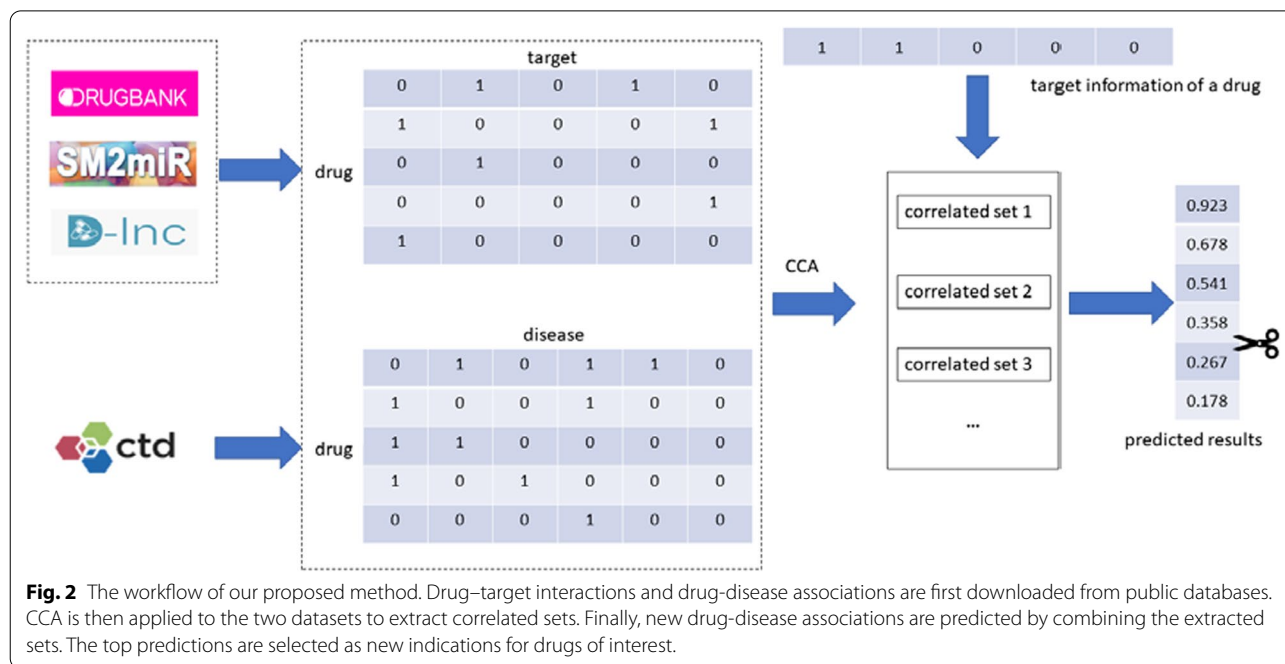
where c_1 and c_2 are parameters to control the sparsity. We refer to this as sparse canonical correlation analysis (SCCA). We apply a strategy of penalized matrix decomposition (PMD) [56] to the matrix $Z = X^T Y$ to obtain the weight vectors α and β .

To receive multiple canonical variates, we use a deflation manipulation iteratively as follows:

$$Z^{k+1} = Z^k - d_k \alpha_k \beta_k^T \tag{4}$$

where α_k and β_k are the weight vectors, and d_k is the singular value obtained in each iteration step. We choose targets and diseases in the k pairs of weight vectors with the highest values as correlated sets.

To predict new indications for a drug with a known target vector x_{new} , we compute the scores of y_{new} by combining the k pairs of weight vectors according to the following equation:



$$y_{new} = \sum_{i=1}^k \beta_i \rho_i \alpha_i^T x_{new} \tag{5}$$

The elements in y_{new} with the highest scores are chosen as the predicted indications for the drug. This prediction strategy was used in previous studies [53, 54]. The workflow of our method is depicted in Fig. 2.

Evaluation metrics

In order to test the prediction performance of our method, we implement 10-fold cross-validations on the drugs. We split the whole drugs into 10 subsets of roughly equal sizes, and each subset is used in turn as a test set. We train our method on the remaining 9 subsets. We prioritize the inferred drug-disease associations according to the prediction scores (see Eq. (5)). Setting different thresholds, true positive rate (TPR) and false positive rate (FPR) are calculated to plot ROC curves. Area under ROC curve (AUC) values are computed for performance evaluation. To obtain robust results, we repeated the cross-validation experiments 10 times.

Moreover, we comprehensively predict novel drug-disease associations for drug repositioning for the drugs not included in the benchmark datasets. We analyse the top-ranked results by searching evidence from the public database CTD [32]. Note we only choose curated records of drug indications in this database for prediction confirmation.

Abbreviations

ncRNA: Non-coding RNAs; CCA: Canonical correlation analysis; MoA: Mechanism of actions; ADR: Adverse drug reactions; PMD: Penalized matrix decomposition.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12920-022-01203-1>.

- Additional file 1.** Distribution of numbers of targets and indications of the 1190 drugs.
- Additional file 2.** The extracted 400 correlated sets by SCCA.
- Additional file 3.** Confirmed top-ranking target-disease associations in component #3.
- Additional file 4.** Confirmed top-ranking target-disease associations in component #6.
- Additional file 5.** Confirmed top-ranking target-disease associations in component #7.
- Additional file 6.** The top 50 predicted indications for the 789 drugs.
- Additional file 7.** The verified drug-disease associations in the top 50 predictions.
- Additional file 8.** The source code and data sets used in this study.

Acknowledgements

Not applicable.

Authors' contributions

HC and JZ performed data preparation. HC and ZZ conceived and designed the experiments. HC and ZZ performed all computational experiments. HC and JZ analyzed the results. HC wrote the paper. All authors read and approved the final manuscript.

Funding

This work was supported by the National Natural Science Foundation of China under Grant Numbers 61862026 and 62172140.

Availability of data and materials

All data used in this study are available from the DrugBank/SM2miR/D-Inc/repoDB/CTD/DisGeNET/HMDD/LncRNADisease databases. The original publications of these databases have been cited in our manuscript. The links for these databases are as follows. The source code and data sets used in this study are available at Additional file 8. DrugBank: <https://go.drugbank.com/>. SM2miR: <http://www.jianglab.cn/SM2miR/>. D-Inc: <http://www.jianglab.cn/D-Inc/>. repoDB: <http://apps.chiragjggroup.org/repoDB>. CTD: <http://ctdbase.org/>. DisGeNET: <http://www.disgenet.org>. HMDD: <http://www.cuilab.cn/hmdd>. LncRNADisease: <http://www.cuilab.cn/lncrnadisease>

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹School of Software, East China Jiaotong University, Nanchang 330013, China.

²School of Computer Science and Engineering, Central South University, Changsha 410083, China. ³School of Computer and Data Science, Henan University of Urban Construction, Pingdingshan 467000, China.

Received: 18 September 2021 Accepted: 2 March 2022

Published online: 06 March 2022

References

1. Strebhardt K, Ullrich A. Paul Ehrlich's magic bullet concept: 100 years of progress. *Nat Rev Cancer*. 2008;8(6):473–80.
2. Overington JP, Al-Lazikani B, Hopkins AL. How many drug targets are there? *Nat Rev Drug Discov*. 2006;5(12):993–6.
3. Peterson RT. Chemical biology and the limits of reductionism. *Nat Chem Biol*. 2008;4(11):635–8.
4. Nobeli I, Favia AD, Thornton JM. Protein promiscuity and its implications for biotechnology. *Nat Biotechnol*. 2009;27(2):157–67.
5. Ashburn TT, Thor KB. Drug repositioning: identifying and developing new uses for existing drugs. *Nat Rev Drug Discov*. 2004;3(8):673–83.
6. Skrott Z, Mistrik M, Andersen KK, Friis S, Majera D, Gursky J, Ozdian T, Bartkova J, Turi Z, Moudry P, et al. Alcohol-abuse drug disulfiram targets cancer via p97 segregase adaptor NPL4. *Nature*. 2017;552(7684):194–9.
7. Lounkine E, Keiser MJ, Whitebread S, Mikhailov D, Hamon J, Jenkins JL, Lavan P, Weber E, Doak AK, Côté S, et al. Large-scale prediction and testing of drug activity on side-effect targets. *Nature*. 2012;486(7403):361–7.
8. Hopkins AL, Groom CR. The druggable genome. *Nat Rev Drug Discov*. 2002;1(9):727–30.
9. Campillos M, Kuhn M, Gavin AC, Jensen LJ, Bork P. Drug target identification using side-effect similarity. *Science*. 2008;321(5886):263–6.
10. Yamanishi Y, Araki M, Gutteridge A, Honda W, Kanehisa M. Prediction of drug–target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics*. 2008;24(13):i232–40.
11. Keiser MJ, Setola V, Irwin JJ, Laggner C, Abbas AI, Hufeisen SJ, Jensen NH, Kujjier MB, Matos RC, Tran TB, et al. Predicting new molecular targets for known drugs. *Nature*. 2009;462(7270):175–81.
12. Chen H, Zhang Z. A semi-supervised method for drug–target interaction prediction with consistency in networks. *PLoS ONE*. 2013;8(5):e62975.
13. Smith LK, Shah RR, Cidlowski JA. Glucocorticoids modulate MicroRNA expression and processing during lymphocyte apoptosis. *J Biol Chem*. 2010;285(47):36698–708.
14. Guo H, Liu J, Ben Q, Qu Y, Li M, Wang Y, Chen W, Zhang J. The aspirin-induced long non-coding RNA OLA1P2 blocks phosphorylated STAT3 homodimer formation. *Genome Biol*. 2016. <https://doi.org/10.1186/s13059-016-0892-5>.
15. Huang Z, Shi J, Gao Y, Cui C, Zhang S, Li J, Zhou Y, Cui Q. HMDD v3.0: a database for experimentally supported human microRNA-disease associations. *Nucleic Acids Res*. 2019;47(D1):D1013–7.
16. Bao Z, Yang Z, Huang Z, Zhou Y, Cui Q, Dong D. LncRNADisease 2.0: an updated database of long non-coding RNA-associated diseases. *Nucleic Acids Res*. 2019;47(D1):D1034–7.
17. Chen H, Zhang Z, Li G. Relating disease-gene interaction network with disease-associated ncRNAs. *IEEE Access*. 2019;7:133521–8.
18. Zhang S, Chen L, Jung EJ, Calin GA. Targeting MicroRNAs with small molecules: from dream to reality. *Clin Pharmacol Ther*. 2010;87(6):754–8.
19. Ling H, Fabbri M, Calin GA. MicroRNAs and other non-coding RNAs as targets for anticancer drug development. *Nat Rev Drug Discov*. 2013;12(11):847–65.
20. Chen H, Zhang Z. A miRNA-driven inference model to construct potential drug-disease associations for drug repositioning. *Biomed Res Int*. 2015;2015:1–9.
21. Chen H, Zhang Z, Peng W. miRDDCR: a miRNA-based method to comprehensively infer drug-disease causal relationships. *SCI REP-UK*. 2017;7(1):1–9.
22. Chen H, Zhang Z. Prediction of drug-disease associations for drug repositioning through drug–miRNA-Disease heterogeneous network. *IEEE Access*. 2018;6:45281–7.
23. Matsui M, Corey DR. Non-coding RNAs as drug targets. *Nat Rev Drug Discov*. 2017;16(3):167–79.
24. Chen H, Cheng F, Li J. iDrug: integration of drug repositioning and drug–target prediction via cross-network embedding. *PLoS Comput Biol*. 2020;16(7):e1008040.
25. Li J, Zheng S, Chen B, Butte AJ, Swamidass SJ, Lu Z. A survey of current trends in computational drug repositioning. *Brief Bioinform*. 2016;17(1):2–12.
26. Piñero J, Ramírez-Anguita JM, Saüch-Pitarch J, Ronzano F, Centeno E, Sanz F, Furlong LI. The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res*. 2020;48:D845–55.
27. Cheng F, Liu C, Jiang J, Lu W, Li W, Liu G, Zhou W, Huang J, Tang Y, Altman RB. Prediction of drug–target interactions and drug repositioning via network-based inference. *PLoS Comput Biol*. 2012;8(5):e1002503.
28. Wu Z, Cheng F, Li J, Li W, Liu G, Tang Y. SDTNBI: an integrated network and cheminformatics tool for systematic prediction of drug–target interactions and drug repositioning. *Brief Bioinform*. 2016. <https://doi.org/10.1093/bib/bbw012>.
29. Zhang W, Liu F, Luo L, Zhang J. Predicting drug side effects by multi-label learning and ensemble learning. *BMC Bioinform*. 2015;16(1):1–11.
30. Deng Y, Xu X, Qiu Y, Xia J, Zhang W, Liu S, Xu J. A multimodal deep learning framework for predicting drug–drug interaction events. *Bioinformatics (Oxford, England)*. 2020;36:4316–22.
31. Chen H, Zhang Z, Zhang J. *In silico* drug repositioning based on the integration of chemical, genomic and pharmacological spaces. *BMC Bioinform*. 2021;22(1):1–12.
32. Davis AP, Grondin CJ, Johnson RJ, Sciaky D, Wieggers J, Wieggers TC, Matingly CJ. Comparative toxicogenomics database (CTD): update 2021. *Nucleic Acids Res*. 2021;49:D1138–43.
33. Luo H, Wang J, Li M, Luo J, Peng X, Wu F, Pan Y. Drug repositioning based on comprehensive similarity measures and Bi-Random walk algorithm. *Bioinformatics*. 2016;32(17):2664–71.
34. Zhang W, Yue X, Lin W, Wu W, Liu R, Huang F, Liu F. Predicting drug-disease associations by using similarity constrained matrix factorization. *BMC Bioinform*. 2018;19(1):1–12.
35. Zeng X, Zhu S, Liu X, Zhou Y, Nussinov R, Cheng F. deepDR: a network-based deep learning approach to in silico drug repositioning. *Bioinformatics*. 2019;35(24):5191–8.
36. Yang M, Luo H, Li Y, Wang J. Drug repositioning based on bounded nuclear norm regularization. *Bioinformatics*. 2019;35(14):i455–63.
37. Gottlieb A, Stein GY, Ruppin E, Sharan R. PREDICT: a method for inferring novel drug indications with application to personalized medicine. *Mol Syst Biol*. 2011;7(1):496.

38. Zhang P, Agarwal P, Obradovic Z. Computational drug repositioning by ranking and integrating multiple data sources. In: Joint European conference on machine learning and knowledge discovery in databases. Berlin, Heidelberg: Springer; 2013, pp. 579–594.
39. Wen Y, Song X, Yan B, Yang X, Wu L, Leng D, He S, Bo X. Multi-dimensional data integration algorithm based on random walk with restart. *BMC Bioinform.* 2021;22:1–22.
40. Liu H, Song Y, Guan J, Luo L, Zhuang Z. Inferring new indications for approved drugs via random walk on drug-disease heterogeneous networks. *BMC Bioinform.* 2016;17(S17):269–77.
41. He S, Wen Y, Yang X, Liu Z, Song X, Huang X, Bo X. PIMD: an integrative approach for drug repositioning using multiple characterization fusion. *Genom Proteom Bioinform.* 2020;18:565–81.
42. Xie L, He S, Zhang Z, Lin K, Bo X, Yang S, Feng B, Wan K, Yang K, Yang J, et al. Domain-adversarial multi-task framework for novel therapeutic property prediction of compounds. *Bioinformatics.* 2020;36(9):2848–55.
43. Xie L, He S, Song X, Bo X, Zhang Z. Deep learning-based transcriptome data classification for drug–target interaction prediction. *BMC Genom.* 2018;19(S7):93–102.
44. Chou T. Drug combination studies and their synergy quantification using the Chou–Talalay method. *Cancer Res.* 2010;70(2):440–6.
45. Liu H, Zhang W, Zou B, Wang J, Deng Y, Deng L. DrugCombDB: a comprehensive database of drug combinations toward the discovery of combinatorial therapy. *Nucleic Acids Res.* 2020;48:D871–81.
46. Ianevski A, Giri AK, Aittokallio T. SynergyFinder 2.0: visual analytics of multi-drug combination synergies. *Nucleic Acids Res.* 2020;48(W1):W488–93.
47. Wu L, Wen Y, Leng D, Zhang Q, Dai C, Wang Z, Liu Z, Yan B, Zhang Y, Wang J, et al. Machine learning methods, databases and tools for drug combination prediction. *Brief Bioinform.* 2021. <https://doi.org/10.1093/bib/bbab355>.
48. 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.* 2018;46(D1):D1074–82.
49. Liu X, Wang S, Meng F, Wang J, Zhang Y, Dai E, Yu X, Li X, Jiang W. SM2miR: a database of the experimentally validated small molecules' effects on microRNA expression. *Bioinformatics.* 2013;29(3):409–11.
50. Jiang W, Qu Y, Yang Q, Ma X, Meng Q, Xu J, Liu X, Wang S. D-Inc: a comprehensive database and analytical platform to dissect the modification of drugs on lncRNA expression. *RNA Biol.* 2019;16(11):1586–91.
51. Brown AS, Patel CJ. A standard database for drug repositioning. *Sci Data.* 2017;4(1):1–7.
52. Hotelling H. Relations between two sets of variates. *Biometrika.* 1936;28:321–77.
53. Pauwels E, Stoven V, Yamanishi Y. Predicting drug side-effect profiles: a chemical fragment-based approach. *BMC Bioinform.* 2011;12(1):169.
54. Mizutani S, Pauwels E, Stoven V, Goto S, Yamanishi Y. Relating drug–protein interaction network with drug side effects. *Bioinformatics.* 2012;28(18):i522–8.
55. Chen H, Zhang Z, Feng D. Prediction and interpretation of miRNA-disease associations based on miRNA target genes using canonical correlation analysis. *BMC Bioinform.* 2019;20(1):1–8.
56. Witten DM, Tibshirani R, Hastie T. A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics.* 2009;10(3):515–34.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

