

Drug repositioning by prediction of drug's anatomical therapeutic chemical code via network-based inference approaches

Yayuan Peng[†], Manjiong Wang[†], Yixiang Xu[†], Zengrui Wu, Jiye Wang, Chao Zhang, Guixia Liu, Weihua Li, Jian Li and Yun Tang

Corresponding authors: Yun Tang, Shanghai Key Laboratory of New Drug Design, School of Pharmacy, East China University of Science and Technology, 130 Meilong Road, Shanghai 200237, China. Tel.: +86-21-64251052; Fax: +86-21-64251033. E-mail: ytang234@ecust.edu.cn; Jian Li, State Key Laboratory of Bioreactor Engineering, East China University of Science and Technology, 130 Meilong Road, Shanghai 200237, China. Tel./Fax: +86-21-64252584. E-mail: jianli@ecust.edu.cn

[†]These authors contributed equally to this work.

Abstract

Drug discovery and development is a time-consuming and costly process. Therefore, drug repositioning has become an effective approach to address the issues by identifying new therapeutic or pharmacological actions for existing drugs. The drug's anatomical therapeutic chemical (ATC) code is a hierarchical classification system categorized as five levels according to the organs or systems that drugs act and the pharmacology, therapeutic and chemical properties of drugs. The 2nd-, 3rd- and 4th-level ATC codes reserved the therapeutic and pharmacological information of drugs. With the hypothesis that drugs with similar structures or targets would possess similar ATC codes, we exploited a network-based approach to predict the 2nd-, 3rd- and 4th-level ATC codes by constructing substructure drug-ATC (SD-ATC), target drug-ATC (TD-ATC) and Substructure&Target drug-ATC (STD-ATC) networks. After 10-fold cross validation and two external validations, the STD-ATC models outperformed the SD-ATC and TD-ATC ones. Furthermore, with KR as fingerprint, the STD-ATC model was identified as the optimal model with AUC values at 0.899 ± 0.015 , 0.916 and 0.893 for 10-fold cross validation, external validation set 1 and external validation set 2, respectively. To illustrate the predictive capability of the STD-ATC model with KR fingerprint, as a case study, we predicted 25 FDA-approved drugs (22 drugs were actually purchased) to have potential activities on heart failure using that model. Experiments *in vitro* confirmed that 8 of the 22 old drugs have shown mild to potent cardioprotective activities on both hypoxia model and oxygen–glucose deprivation model, which demonstrated that our STD-ATC prediction model would be an effective tool for drug repositioning.

Key words: anatomical therapeutic chemical code; drug repositioning; network-based inference

Yayuan Peng and Jiye Wang are PhD students at East China University of Science and Technology, Shanghai, China. They are working on bioinformatics and multi-omics analysis.

Manjiong Wang, Yixiang Xu and Chao Zhang are PhD students at East China University of Science and Technology, Shanghai, China. They are interested in drug design for heart failure and Alzheimer's disease.

Zengrui Wu is a post-doc fellow at East China University of Science and Technology, Shanghai, China. His current research interests include system pharmacology and drug repositioning.

Jian Li is the professor at East China University of Science and Technology, Shanghai, China. His research group has been working on drug design and optimization for serious diseases, including heart failure, Alzheimer's disease, diabetes and so on.

Yun Tang, Guixia Liu and Weihua Li are professors at East China University of Science and Technology, Shanghai, China. Their research interests focus on computer-aided drug design, molecular modeling, bioinformatics and artificial intelligence.

Submitted: 2 January 2020; Received (in revised form): 5 February 2020

Introduction

Drug discovery and development is a difficult process with high failure rate and long time-consuming. The probability of a drug from phase I to launch stays statistically at less than 10% [1]. Therefore, drug repositioning has been one of the primary methods to address this dilemma by identifying new uses for old drugs [2–4].

The anatomical therapeutic chemical (ATC) code is proposed by the World Health Organization (WHO) for international drug utilization studies since 1981. It is a hierarchical classification system categorized as five different levels according to the organs or systems on which drugs act and according to the pharmacology, therapeutic and chemical properties of drugs [5]. This system has 14 main anatomical or pharmacological groups (i.e. the 1st-level code) as shown in Table 1. The 1st-level code could be further divided into therapeutic groups (i.e. the 2nd-level code). Later, the 3rd- and 4th-level codes stand for pharmacological/therapeutic subgroups and chemical characteristic, and the 5th-level code represents chemical feature merely. Therefore, the 2nd-, 3rd- and 4th-level ATC codes are often used to identify the therapeutic and pharmacological information of one drug. As to a chemical substance, if their 2nd-, 3rd- and 4th-level ATC codes could be defined, the potential therapeutic and pharmacological actions would be distinct to researchers. Taking aspirin as an example, it has B01AC06 and N02BA01 two quite different ATC codes. The B01AC06 stands for the antithrombotic therapy, whereas N02BA01 represents the anti-fever effect. The detailed ATC code levels of aspirin were listed in Table 2 [6]. Drugs could possess multiple ATC codes due to their different anatomical therapeutic and pharmacological actions.

Up to now, only 2119 of 9950 small molecules have known ATC codes according to the statistical data of DrugBank (<https://www.drugbank.ca/>, version 5.1.2) [7]. At the same time, the ATC codes of the 2119 drugs obey power law distribution according to our statistics (Figure 1). In other words, most of drugs only have one ATC code, whereas some drugs have two or more ATC codes. Up to May 2019, as the world's largest database of freely accessible chemical information, PubChem includes about 98 million compounds [8], and the number increases very quickly year by year thanks to more advanced synthetic technologies.

Table 1. The 14 main anatomical or pharmacological groups (1st level) of ATC code

1st-level code	Anatomical or pharmacological groups
A	Alimentary tract and metabolism
B	Blood and blood-forming organs
C	Cardiovascular system
D	Dermatologicals
G	Genitourinary system and sex hormones
H	Systemic hormonal preparations, excl. Sex hormones and insulins
J	Anti-infectives for systemic use
L	Antineoplastic and immunomodulating agents
M	Musculoskeletal system
N	Nervous system
P	Antiparasitic products, insecticides and repellents
R	Respiratory system
S	Sensory organs
V	Various

Thus, the imbalanced data between limited drugs with known ATC codes and plentiful compounds without known pharmacological property make it necessary to develop new methods for the prediction of therapeutic and pharmacological function of those undefined molecules.

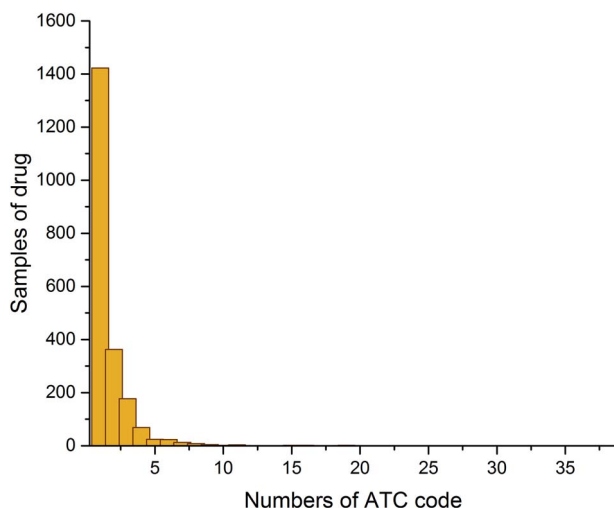
For a long time, the 'one drug → one target → one disease' paradigm was popular in the process of drug discovery and development. However, many effective drugs could act on multiple proteins, namely, polypharmacology, such as metformin and aspirin [9, 10]. Metformin has been found to treat influenza in 1950 and reduce cardiovascular risk in 2008, respectively, besides glucose-lowering effect [11, 12]. The miracle drug, aspirin, was used as an antipyretic and anti-inflammatory agent originally and then applied to prevent cardiovascular and cerebrovascular diseases [13]. Aspirin has been proven to prevent colorectal and other types of cancer later [14]. Therefore, network pharmacology was born at the right moment with 'multiple drugs → multiple targets → multiple diseases' mode and has been applied in drug discovery and repositioning [15]. When it comes to drug repositioning, one indirect approach is target-based, which finds and validates a target with specific verified therapeutic effects for old drugs [16, 17]. Another approach is phenotype-based, i.e. to find the direct effects between molecules and phenotypes. Compared with target-based approaches, phenotype-based ones could jump over the tedious target validation procedure and consider the direct therapeutic or pharmacologic effects, which is time and experiment saving [18].

There are several computational methods reported for the prediction of ATC code. As early as in 2008, Dunkel et al. [19] developed a webserver named SuperPred for drug-ATC code and target prediction. The main principle of SuperPred is that drugs with similar structures will possess similar biological activities. The authors declared that a similarity score higher than 0.85 would lead to correct ATC prediction for 81% of all cases [19]. In 2013, Wang et al. [20] applied a machine learning method, support vector machine, to predict connections of drugs and ATC codes. The data set they used was from the work of Yamanishi et al. [21]. They found that about 75% drug pairs would share common ATC codes if both their chemical structure similarity and target protein similarity scores were larger than 0.8. In 2015, Chen et al. [22] proposed dD-hybrid to predict ATC codes by incorporating drug-domain interaction network information into prediction models. They discovered that drugs connected with common domains would have the same ATC codes. In 2017, Olson et al. [6] compared four different machine learning methods in the tiered learning architecture to predict the first four-level ATC codes, including multilayer perceptron, naïve Bayes, random forest and support vector machine. In 2018, Chen et al. [23] utilized the shortest path and random walk methods to infer the 14 main classes at the 1st-level ATC code classification system. In 2019, Wang et al. [24] adopted network-based label space partition (NLSP) method to the prediction of ATC code within the multilabel learning framework, which was based on structural and fingerprint similarities of a compound to others belonging to different ATC categories. However, all the abovementioned methods have no experimental validations.

With the hypothesis that drugs with similar structures or protein targets would possess similar pharmacological and therapeutic effects [19], we could infer new effects for FDA-approved drugs and new chemical entities based on the ATC code classification system. We hence proposed a network-based inference approach to predict the 2nd-, 3rd- and 4th-level ATC codes in this study. Figure S1 (see [Supplementary Data](#) available online at <http://bib.oxfordjournals.org/>) showed the whole workflow

Table 2. ATC codes of Aspirin: B01AC06 and N02BA01

ATC level	ATC codes of aspirin	
	Antiplatelet effects	Anti-fever/pain reliever
1st: Anatomical main group	B: Blood and blood-forming agents	N: Nervous system
2nd: Therapeutic main group	B01: Antithrombotic agents	N02: Analgesic
3rd: Therapeutic/pharmacological subgroup	B01A: Antithrombotic agents	N02B: Other analgesics and antipyretics (fever reducers)
4th: Chemical/therapeutic/pharmacological subgroup	B01AC: Platelet aggregation inhibitors	N02BA: Salicylic acid and derivatives
5th: Chemical substance identifiers	B01AC06: Acetylsalicylic acid	N02BA01: Acetylsalicylic acid

**Figure 1.** Histogram of the 2119 drugs with known ATC codes in DrugBank, which has a power law distribution.

of this study. Specifically, we built substructure drug-ATC (SD-ATC), target drug-ATC (TD-ATC) and ubstructure&Target drug-ATC (STD-ATC) tripartite networks and validated the models via 10-fold validation and two external validation sets. Based on known networks, we could predict the 2nd-, 3rd- and 4th- level ATC codes for FDA-approved drugs and new chemicals. As an example to apply the method, we utilized the best model STD-ATC with KR as the fingerprint to predict old drugs with potential activities on heart failure. The prediction was further verified through *in vitro* experiments.

Material and methods

Data collection and preparation

Construction of drug-ATC code association network

The ATC code annotations of drugs, together with their DrugBank IDs and SMILES strings, were extracted from DrugBank (<https://www.drugbank.ca/>, version 5.1.2) [7]. A number of C atoms and canonical SMILES strings were calculated by Open Babel toolkit (version 2.4.1) [25]. Afterwards, drugs with less than three C atoms were removed. The associations of drugs and ATC codes were then formed a bipartite network. Here the ATC code was represented in the 2nd, 3rd and 4th levels. For instance, aspirin has two ATC codes: B01AC06 and N02BA01. Then the associations between aspirin and ATC codes were represented as six pairs, including aspirin-B01, aspirin-B01A, aspirin-B01AC, aspirin-N02, aspirin-N02B and aspirin-N02BA.

Construction of drug-substructure association network

In this study, seven types of fingerprints were used to characterize the substructures of compounds, including CDK fingerprint (CDK), CDK extended fingerprint (CDKExt), substructure fingerprint (FP4), CDK graph-only fingerprint (Graph), Klekota-Roth fingerprint (KR), MACCS fingerprint (MACCS) and PubChem fingerprint (PubChem). Those fingerprints adopted diverse bits to describe the substructures of small molecules. PaDEL-Descriptor (version 2.2.1) was used to calculate the said fingerprints for each molecule [26]. Similarly, the associations of drugs or new chemicals and substructures were also represented as a bipartite network.

Construction of drug-target interaction network

The interactions of drugs and experimentally validated protein targets were gathered from four databases, namely, ChEMBL [27], BindingDB [28], IUPHAR/BPS Guide to PHARMACOLOGY [29] and PDSP K_i Database [30]. Drug-target interactions were preserved when the following criteria were satisfied: (i) targets belong to *Homo sapiens* rather than the mussy; (ii) half maximal inhibitory concentration (IC₅₀), inhibitory constant (K_i), dissociation constant (K_d) or effective concentration (EC₅₀) less than 10 μM or potency and AC₅₀ less than 10 μM were labeled as active, where the binding criteria were adopted in some researches [31]; (iii) redundant targets were removed. In this way the drug-target interaction data were gathered. The drug-target interactions were also represented as a binary network.

Construction of the tripartite contextual network

After the drug-ATC code, drug-substructure, and drug-target networks were prepared respectively, we constructed SD-ATC, TD-ATC and STD-ATC three tripartite networks, respectively, by connecting the abovementioned bipartite networks. Particularly, STD-ATC networks were constructed by combining the corresponding drug-substructure network and drug-target network and then connecting with drug-ATC code network. The drug was the intermediate link among substructure, target or Substructure&Target and ATC code. DrugBank ID number was used as the uniform identifier of each drug. The schematic diagram of the tripartite networks was shown in Figure 2.

We denoted that $D = \{D_1, D_2, \dots, D_{N_D}\}$ is a node set of N_D known drugs, $S = \{S_1, S_2, \dots, S_{N_S}\}$ is a node set of N_S substructures, $T = \{T_1, T_2, \dots, T_{N_T}\}$ is a set of N_T targets, $U = S \cup T$ is a union node set of N_S substructures and N_T targets, and $A = \{A_1, A_2, \dots, A_{N_A}\}$ is a node set of N_A ATC codes. Then the tripartite networks can be represented as tripartite graphs $G_1(V_1, E_1)$, $G_2(V_2, E_2)$ and $G_3(V_3, E_3)$, where V is the set of vertices and E is the set of edges containing drug-ATC pairs and drug-substructure pairs,

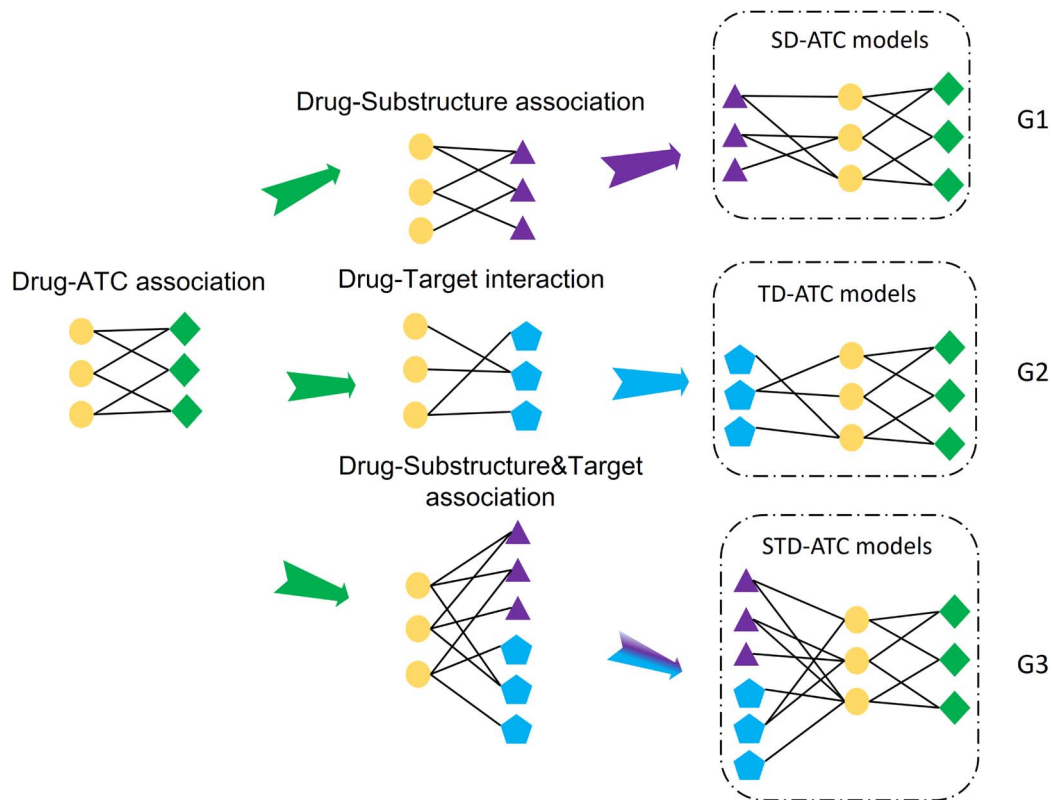


Figure 2. The schematic diagram of the three types of tripartite networks, namely, SD-ATC, TD-ATC and STD-ATC.

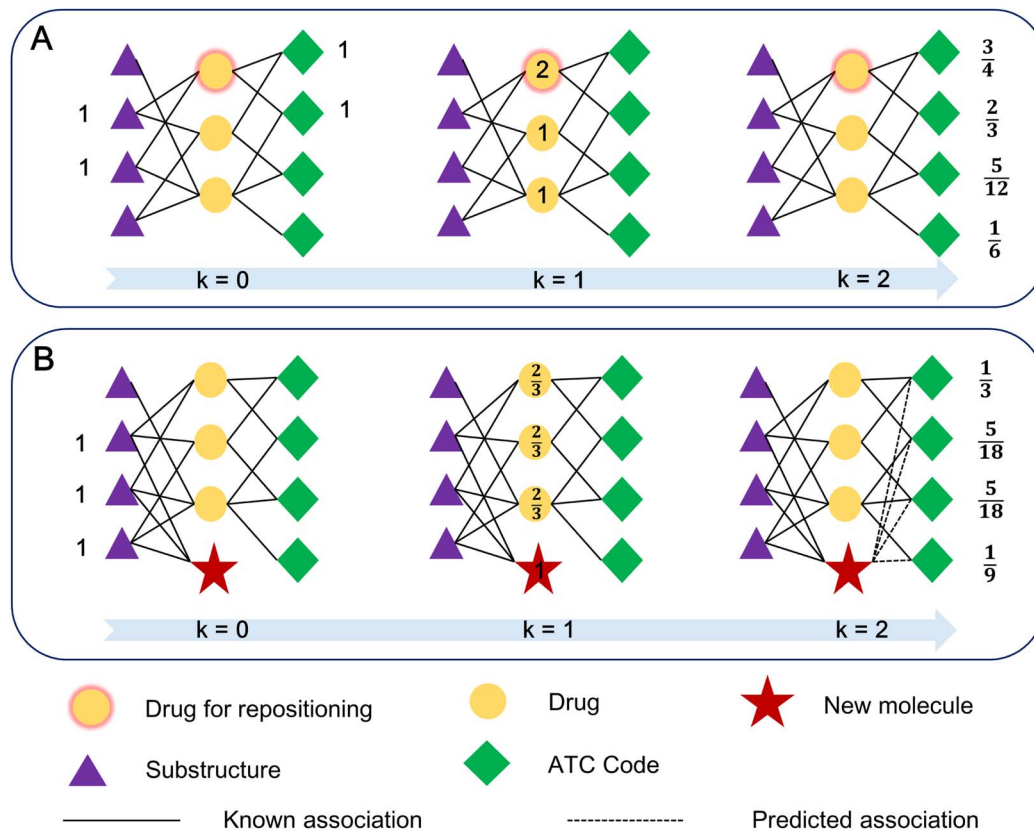


Figure 3. The schematic diagram of resource diffusion process in SD-ATC network. New associations between drugs or new molecules and ATC codes can be predicted. The predicted results can be ranked according to the final resource value located in nodes of ATC code. (A) By resource diffusion, potential new ATC codes can be recommended for drug repositioning. (B) As to a new molecule, the associations between new molecule and ATC codes can be predicted for phenotype-based virtual screening.

drug-target or drug-Substructure&Target pairs. Specifically, besides they all have the same vertices of drugs and ATC codes, V1 has vertices of substructures, V2 has vertices of targets and V3 has vertices of substructures and targets. Accordingly, in addition to edges of drug-ATC pairs, E1 has edges of drug-substructure pairs, E2 has edges of drug-target pairs and E3 has edges of drug-Substructure&Target pairs. Then the tripartite network is capable to predict new ATC codes for drugs.

Description of the network-based inference method

The principle of network-based inference method is that items with similar properties will share similar objects. The calculation involves a process of resource diffusion. In this study, for each drug D_i in the network, it has initial resource located in its neighbor connected nodes. Take G1 as an example (Figure 3A), for drug D_i , its neighbor nodes connected with it will obtain an equal initial resource, including substructure nodes and ATC codes ($k=0$). Then the nodes of substructures and ATC codes will diffuse their resource to their neighbor nodes equally back, namely, nodes of drugs ($k=1$). Later the nodes of drugs will send the resource to nodes of substructure and ATC codes again ($k=2$). Such two steps of resource diffusion could continue everlastingly. It is noteworthy that only the value of k is even ($k=2, 4, 6, 8, 10$, etc.), the final resource can reach to nodes of ATC codes. Via such kind of resource diffusion, a drug to be predicted (drug A) can get new ATC codes for drug repositioning. The final resource values located in different ATC codes stand for the rank of corresponding ATC codes. For each drug to be predicted for ATC codes, the larger is the final resource value in certain ATC code, the higher is the rank among other else ATC codes. A higher rank means a bigger possibility for drug A to have such an ATC code.

Mathematically, graphs G1, G2 and G3 can be represented as adjacency matrix W1, W2 and W3, which is a $(N_D + N_S + N_A)$, $(N_D + N_T + N_A)$ and $(N_D + N_U + N_A)$ order square matrix, respectively:

$$W1 = \begin{bmatrix} O & M_{DS} & M_{DA} \\ M_{DS}^T & O & O \\ M_{DA}^T & O & O \end{bmatrix} \quad (1)$$

$$W2 = \begin{bmatrix} O & M_{DT} & M_{DA} \\ M_{DT}^T & O & O \\ M_{DA}^T & O & O \end{bmatrix} \quad (2)$$

$$W3 = \begin{bmatrix} O & M_{DU} & M_{DA} \\ M_{DU}^T & O & O \\ M_{DA}^T & O & O \end{bmatrix} \quad (3)$$

where M_{DS} is a $N_D \times N_S$ matrix, M_{DT} is a $N_D \times N_T$ matrix, M_{DU} is a $N_D \times N_U$ matrix and M_{DA} is a $M_D \times M_A$ matrix, whereas M_{DS}^T , M_{DT}^T , M_{DU}^T and M_{DA}^T are the transposed matrix of M_{DS} , M_{DT} , M_{DU} and M_{DA} , respectively. They can be defined as follows:

$$M_{DS}(i, j) = \begin{cases} 1 & \text{if } D_i \text{ is linked with } S_j \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

$$M_{DT}(i, j) = \begin{cases} 1 & \text{if } D_i \text{ is linked with } T_j \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

$$M_{DU}(i, j) = \begin{cases} 1 & \text{if } D_i \text{ is linked with } U_j \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

$$M_{DA}(i, j) = \begin{cases} 1 & \text{if } D_i \text{ is linked with } A_j \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

Let R1 be a $(N_D + N_S + N_A)$, R2 be a $(N_D + N_T + N_A)$ and R3 be a $(N_D + N_U + N_A)$ order square matrix, respectively, defined as below:

$$R1(i, j) = \frac{W1(i, j)}{\sum_{l=1}^{N_D+N_S+N_A} w1(i, l)} \quad (8)$$

$$R2(i, j) = \frac{W2(i, j)}{\sum_{l=1}^{N_D+N_T+N_A} w2(i, l)} \quad (9)$$

$$R3(i, j) = \frac{W3(i, j)}{\sum_{l=1}^{N_D+N_U+N_A} w3(i, l)} \quad (10)$$

The final resource matrix can be calculated as below:

$$Y1 = W1 \times R1^k \quad (11)$$

or

$$Y2 = W2 \times R2^k \quad (12)$$

or

$$Y3 = W3 \times R3^k \quad (13)$$

where $R1^k$, $R2^k$ and $R3^k$ are transfer matrix and k is the number of resource-spreading process. The value of $Y1(i, N_D + N_S + j)$, $Y2(i, N_D + N_T + j)$ or $Y3(i, N_D + N_U + j)$ is the score of D_i-A_j interaction.

Through the abovementioned matrix operation, the potential ATC codes can be recommended for a drug by resource diffusion process in the SD-ATC network, TD-ATC network or STD-ATC network. Namely, the prediction process was conducted by G1, G2 or G3. The prediction performance of the aforesaid three kinds of conditions, substructure, target and Substructure&Target, would be compared to obtain the optimal model.

Evaluation of prediction models

The prediction models could be divided into three categories, namely, SD-ATC, TD-ATC and STD-ATC. SD-ATC models mean that fingerprints were used to predict ATC codes. DT-ATC model stands for the prediction of ATC codes through protein target information. SDT-ATC models mean that both substructures and targets were used to predict ATC codes simultaneously. Seven different fingerprints were calculated to characterize the substructures of drugs or compounds. Therefore, SD-ATC models had CDK, CDKExt, FP4, Graph, KR, MACCS and PubChem models for short, TD-ATC model had Target model for short, while STD-ATC models consisted of CDK&Target, CDKExt&Target, FP4&Target, Graph&Target, KR&Target, MACCS&Target and PubChem&Target models. Totally, 15 models were built for the prediction of ATC codes, which were summarized in Table S1 (see Supplementary Data available online at <http://bib.oxfordjournals.org/>).

10-fold cross validation

The 10-fold cross validation, as a kind of approach to verify the performance of prediction models, was widely used in *in silico* target prediction studies [16, 32]. The principle of 10-fold cross validation is straightway. In our study, for the different models to be evaluated, the drug-ATC associations were divided into 10 parts randomly and equally. One part was served as the test set, while the remaining nine parts were used as the training set to build the model and predict the test set. This process was

repeated for 10 times until all the 10 parts were predicted for only once. Then, the evaluation indicators were calculated. To reduce randomness, the 10-fold cross validation was repeated by 10 times so that 100 groups of evaluation indicators can be obtained to evaluate the prediction model.

External validation

The actual generalization abilities of 15 prediction models were also evaluated by two external validation sets. In external validation sets, drugs with known ATC codes were seen as new molecules. That is to say, for drugs in a validation set, their ATC codes were reassigned by different prediction models. Then the evaluation indicators were calculated to measure the indexes of the prediction models. In this study, two external validation sets were used to inspect the prediction models. Their known ATC codes were collected from DrugBank. Information of substructures and targets were gathered in the same way as that for training set. Validation set 1 had known targets, whereas validation set 2 did not have known targets. By such a way, for some new molecules without known protein targets, their ATC code prediction accuracy can also be calculated.

Calculation of evaluation indicators

Through comparing the known ATC codes in test set and the predicted ATC codes by training set, the properties of prediction models can be measured by a series of evaluation indicators. Here two indicators were calculated, including precision (P) and recall (R). In our previous study, these two indicators were also used to measure the performance of network-based target prediction model [16, 17, 33–35]. The implication of these two indicators was briefly shown as below:

$$P(L) = \frac{1}{M} \cdot \sum_{i=1}^M \frac{X_i(L)}{L} \quad (14)$$

$$R(L) = \frac{1}{M} \cdot \sum_{i=1}^M \frac{X_i(L)}{X_i} \quad (15)$$

where M and N are the numbers of drugs and ATC codes participated in evaluation. L means the length of ATC codes to be predicted for each drug. X_i is the missing drug-ATC pairs for drug D_i . $X_i(L)$ stands for the number of the true-positive prediction ranked in top L of the predicted ATC codes for drug D_i (i.e. missing drug-ATC pairs were correctly predicted for drug D_i). X represents the total number of missing drug-ATC pairs. Furthermore, a receiver operating characteristic (ROC) curve was generated by computing a series of true-positive and false-positive rates under different L . The areas under these ROC curves, namely, AUC, were calculated to further describe the performance of prediction models.

Case study: drug repositioning for heart failure

Heart failure (HF) is one of the most life-threatening cardiovascular diseases for people older than 60 years [36]. The prevalence of HF increases as the average aging around the world increases [37]. Inhibitors of the renin-angiotensin-aldosterone systems and β -blockers were the common medicines to HF. However, all approved drugs for confrontation heart failure were used for the prevention and delay of HF but not cure. Therefore, it is crucial to discover new potent drugs for the treatment of HF.

Computational prediction

Data collection of approved drugs for heart failure. To collect known ATC code data of HF therapy, 46 approved drugs for HF therapy were obtained from DrugBank, among which 43 drugs have assigned ATC codes. The known ATC codes of these drugs were extracted and prepared as the 2nd-, 3rd- and 4th-level codes. Totally, 75 HF-related ATC codes were gathered (Table S2, see Supplementary Data available online at <http://bib.oxfordjournals.org/>).

Prediction of ATC codes. After 10-fold cross validation and two external validations, we used the best model to predict ATC codes for approved drugs. Only the approved drugs in DrugBank labeled as 'approved'; 'investigational, approved'; 'experimental, approved'; 'investigational, experimental approved'; 'yet_approved'; 'yet_approved, approved'; 'yet_approved, investigational, approved'; 'yet_approved, experimental'; and 'yet_approved, investigational, approved' were predicted. And the length of prediction is 15. That is, each drug predicted 15 potential ATC codes.

Fisher's exact test. The number of an approved drug's ATC codes related to HF conforms approximately to hypergeometric distribution, and its probability mass function is as follows:

$$P(X = k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}} \quad (16)$$

where N is the total ATC codes in our prediction model, K is the total number of HF-related ATC codes in our model, n is the number of ATC codes of each drug, and k is the number of HF-related ATC codes in all ATC codes for each drug. And $P(X = k)$ stands for the probability of k HF-related ATC codes occurring in predicted ATC codes for each approved drug. Fisher's exact test was implemented to assess the significance of enrichment of HF-related ATC codes in all approved drugs. P -value was calculated and adjusted by Benjamini Hochberg method. Then all approved drugs were ranked based on their adjusted P -values. Drugs with P -value ≤ 0.05 were selected for experimental test.

Experimental validation

We used both hypoxia model and oxygen-glucose deprivation (OGD) model to verify the above predicted results. The following are the detailed methods for *in vitro* experimental validation process.

Cell culture. H9C2 (2–1), a rat cardiomyocyte cell line, was purchased from the Stem Cell Bank, Chinese Academy of Sciences. Cells were grown in Dulbecco's Modified Eagle Medium (DMEM; HyClone #SH30243.01) supplemented with 10% fetal bovine serum (FBS; Gibco #10270-106) and 1% penicillin/streptomycin (YESEN #60162ES76) at 37°C with 5% CO₂ atmosphere.

Hypoxia model. Cells were seeded in 96-well plates (Corning #3599) and divided into control group, hypoxia injury model group and intervention groups of tested drugs. After 12 h, cells were exposed to the tested drugs for 24 h. These drugs were dissolved in dimethyl sulfoxide (DMSO) and diluted to desired concentration with culture medium. Then, cells were subjected to hypoxia treatments. For normoxic group, cells were maintained at normal atmosphere, while for hypoxia injury groups, cells were incubated in a hypoxic incubator at 37°C for 24 h.

LDH release assay. Cell injury was measured by the amount of LDH released into the extracellular fluid. After 24-h hypoxia

Table 3. The performance of different models in the 10-fold cross validation and two external validation sets when $k = 2$ and $L = 15$

		Models	AUC	P	R
Training set	SD-ATC models	CDK	0.828 ± 0.013	0.082 ± 0.012	0.108 ± 0.017
		CDKExt	0.822 ± 0.012	0.080 ± 0.012	0.107 ± 0.017
		FP4	0.848 ± 0.013	0.116 ± 0.017	0.156 ± 0.025
		Graph	0.834 ± 0.013	0.095 ± 0.013	0.125 ± 0.018
		KR	0.885 ± 0.015	0.196 ± 0.024	0.246 ± 0.030
		MACCS	0.824 ± 0.012	0.091 ± 0.013	0.120 ± 0.019
	TD-ATC model	PubChem	0.838 ± 0.013	0.101 ± 0.015	0.133 ± 0.022
		Target	0.839 ± 0.021	0.218 ± 0.024	0.297 ± 0.032
	STD-ATC models	CDK&Target	0.844 ± 0.013	0.098 ± 0.014	0.129 ± 0.021
		CDKExt&Target	0.838 ± 0.013	0.094 ± 0.014	0.125 ± 0.020
		FP4&Target	0.882 ± 0.014	0.207 ± 0.022	0.275 ± 0.030
		Graph&Target	0.856 ± 0.013	0.116 ± 0.016	0.154 ± 0.023
		KR&Target	0.899 ± 0.015	0.224 ± 0.025	0.285 ± 0.032
		MACCS&Target	0.860 ± 0.013	0.139 ± 0.018	0.184 ± 0.025
External validation set 1	SD-ATC models	PubChem&Target	0.858 ± 0.013	0.124 ± 0.018	0.163 ± 0.026
		CDK	0.855	0.060	0.236
		CDKExt	0.849	0.056	0.226
		FP4	0.878	0.080	0.324
		Graph	0.859	0.064	0.262
		KR	0.906	0.111	0.437
	TD-ATC model	MACCS	0.852	0.061	0.253
		PubChem	0.861	0.066	0.261
	STD-ATC models	Target	0.868	0.119	0.473
		CDK&Target	0.871	0.070	0.274
		CDKExt&Target	0.866	0.065	0.264
		FP4&Target	0.907	0.116	0.461
		Graph&Target	0.881	0.080	0.321
		KR&Target	0.916	0.125	0.489
External validation set 2	SD-ATC models	MACCS&Target	0.884	0.087	0.342
		PubChem&Target	0.880	0.079	0.314
		CDK	0.827	0.050	0.199
		CDKExt	0.819	0.046	0.188
		FP4	0.854	0.074	0.326
		Graph	0.837	0.057	0.231
	STD-ATC models	KR	0.891	0.105	0.440
		MACCS	0.822	0.048	0.200
		PubChem	0.847	0.056	0.222
		CDK&Target	0.828	0.051	0.204
		CDKExt&Target	0.821	0.047	0.190
		FP4&Target	0.857	0.082	0.356
	STD-ATC models	Graph&Target	0.839	0.058	0.235
		KR&Target	0.893	0.108	0.450
		MACCS&Target	0.826	0.053	0.217
		PubChem&Target	0.848	0.058	0.228

injury, 50 μ L of supernatant were transferred into the 96-well plates. Then 50 μ L of the CytoTox 96[®] Reagent (Promega #G1780) was added into each well. The plates were incubated at room temperature away from the light for 30 min. Fifty μ L/well of stop solution was added to stop the reaction, and the absorbance was measured at 490 nm by microplate reader (Biotek, Synergy H1).

OGD model. H9C2 (2–1) cells were seeded in 96-well plates, and they were incubated in culture medium containing tested drugs for 6 h prior to OGD treatments. Then the culture medium was replaced with PBS buffer except the control group, and cells in intervention groups were still exposed to tested drugs. Hypoxia injury was induced in a hypoxic incubator for 3 h.

Cell viability assay. Cell viability was assessed using the Cell Counting Kit-8 (CCK-8; APEX-BIO; #K1018). After the OGD treatments, the supernatant was discarded and 100 μ L/well of the

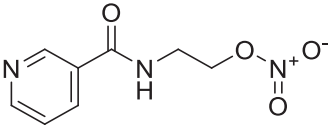
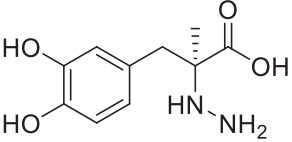
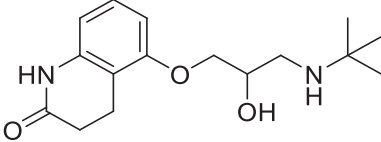
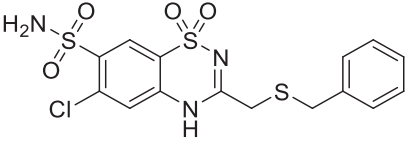
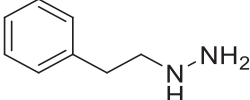
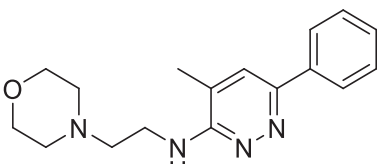
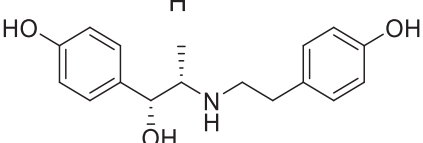
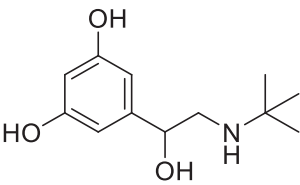
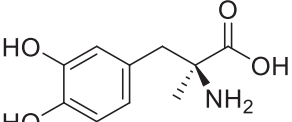
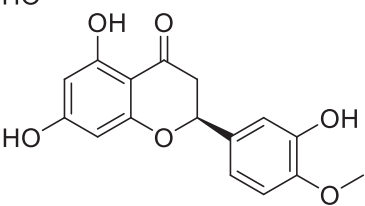
culture medium with 10% CCK-8 was added. The plates were incubated for 2 h at 37°C. The absorption at 450 nm was detected using a microplate reader (Biotek, Synergy H1):

$$\text{Cell viability} = \left[\frac{(\text{OD}_i - \text{OD}_{\text{blank}})}{(\text{OD}_{\text{normal}} - \text{OD}_{\text{blank}})} \right] \times 100\% \quad (17)$$

where OD_i is the absorption of the compounds or model group; $\text{OD}_{\text{normal}}$ is the absorption of the normal group, and OD_{blank} is the absorption of background:

$$\text{Cardioprotective activity} = \left[\left(\text{cell viability}_{\text{compound}} - \text{cell viability}_{\text{model}} \right) / \left(100 - \text{cell viability}_{\text{model}} \right) \right] \times 100\% \quad (18)$$

Table 4. The cell viabilities and cardioprotective activities of the 13 old drugs on OGD model

Num	DrugBank ID	Structure	Cell viability (%)		Cardioprotective activity (%)	
			10 μ M	100 μ M	10 μ M	100 μ M
Positive control	DB09220 (nicorandil)		/	50.39	/	9.32
1	DB00190		50.31	50.08	9.18	8.76
2	DB00521 ^a		46.06	48.74	1.41	6.31
3	DB00562		47.50	44.90	4.04	-0.71
4	DB00780 ^b		47.65	19.59	4.31	-46.97
5	DB00805 ^c		49.21	50.01	7.17	8.63
6	DB00867 ^a		44.66	47.49	-1.15	4.02
7	DB00871 ^d		45.11	50.58	-0.33	9.67
8	DB00968		46.27	57.32	1.79	21.99
9	DB01094		44.42	49.44	-1.59	7.59

Continued

Table 4. Continue

Num	DrugBank ID	Structure	Cell viability (%)		Cardioprotective activity (%)	
			10 μ M	100 μ M	10 μ M	100 μ M
10	DB01193 ^a		44.25	48.40	-1.90	5.68
11	DB02701		45.06	53.14	-0.42	14.35
12	DB06707		45.13	50.86	-0.29	10.18
13	DB11268		49.60	52.47	7.88	13.12

The cell viability of model group is 45.29%

Salt type: ^ahydrochloride, ^bsulfate, ^cdihydrochloride, ^dhemisulfate. “/” indicates no test

Results

Data collection

Totally 15 models were established for the prediction of ATC codes (Table S1, see Supplementary Data available online at <http://bib.oxfordjournals.org/>). In order to evaluate the generalization ability of these models, two external validation sets were prepared. The detailed data were summarized in Table S3 (see Supplementary Data available online at <http://bib.oxfordjournals.org/>).

From DrugBank, we totally obtained 2006 drugs with known ATC codes after data standardization. Among them, 1312 drugs have experimentally validated protein targets in ChEMBL, BindingDB, IUPHAR/BPS Guide to PHARMACOLOGY and PDSP Ki Database, whereas 694 drugs do not have matched targets yet. For the 1312 drugs with targets, we further divided them into training set (1050 drugs) and external validation set 1 (262 drugs) in the ratio of 8:2 randomly. For the 694 drugs without targets, they were considered as external validation set 2.

Specifically, there were 4588 drug-ATC pairs between 1050 drugs and 801 ATC codes in the training set, 1081 drug-ATC pairs between 262 drugs and 448 ATC codes in the external validation set 1 and 2459 drug-ATC pairs between 694 drugs and 594 ATC codes in the external validation set 2. All drugs in the external data sets were excluded in the training set.

Model construction

To evaluate the performance of our models, 10-fold cross validation was applied under different conditions, including different numbers of resource-spreading processes (k), different prediction models and different prediction length (L). As to each model, 10-fold cross validation was repeated for 10 times. After this, the average and SD of evaluation indicators were calculated to measure the performance of each model.

Different resource-spreading processes

For the reason that AUC is independent of prediction length (L), different k values (2, 4, 6, 8, 10, 12, 14, 16) were investigated using AUC under 15 models and $L=5$. From Figure 4A–C, it was found that the AUC values of SD-ATC, TD-ATC and STD-ATC models decreased with the increase of k integrally. When $k=2$, all models, except TD-ATC, reached the optimal AUC because of the lower sparsity of target drug network (Table S1, see Supplementary Data available online at <http://bib.oxfordjournals.org/>). In SD-ATC and STD-ATC models, when $k \geq 10$, AUC decreased slowly and tended to be gentle. Although TD-ATC model obtained the best AUC when $k=4$, its AUC was still lower than the best models of SD-ATC and STD-ATC at $k=2$. Hence, 2 was selected as the best value for k , which was corresponded to the previous network-based target prediction methods [17, 32]. In Figure 4A, KR model and FP4 model achieved better performance. As to STD-ATC models (Figure 4C), the KR&Target model and FP4&Target model also performed better than the other ones.

Variations of prediction length

The values of precision (P) and recall (R) were varied with different prediction length (L). Under the best value of k ($k=2$), the values of P and R were computed when L ranged from 1 to 200. From Figure 4D–F, in each model, the value of P decreased with the increase of L . And in Figure 4G–I, we could find that the value of R increased as L increased. In SD-ATC models, we could find that KR and FP4 models were the top two best. In STD-ATC models, we discovered that KR&Target and FP4&Target models outperformed the others.

The performance of different prediction models

According to the aforementioned analysis, the KR and FP4 models were more splendid than the other SD-ATC models, while

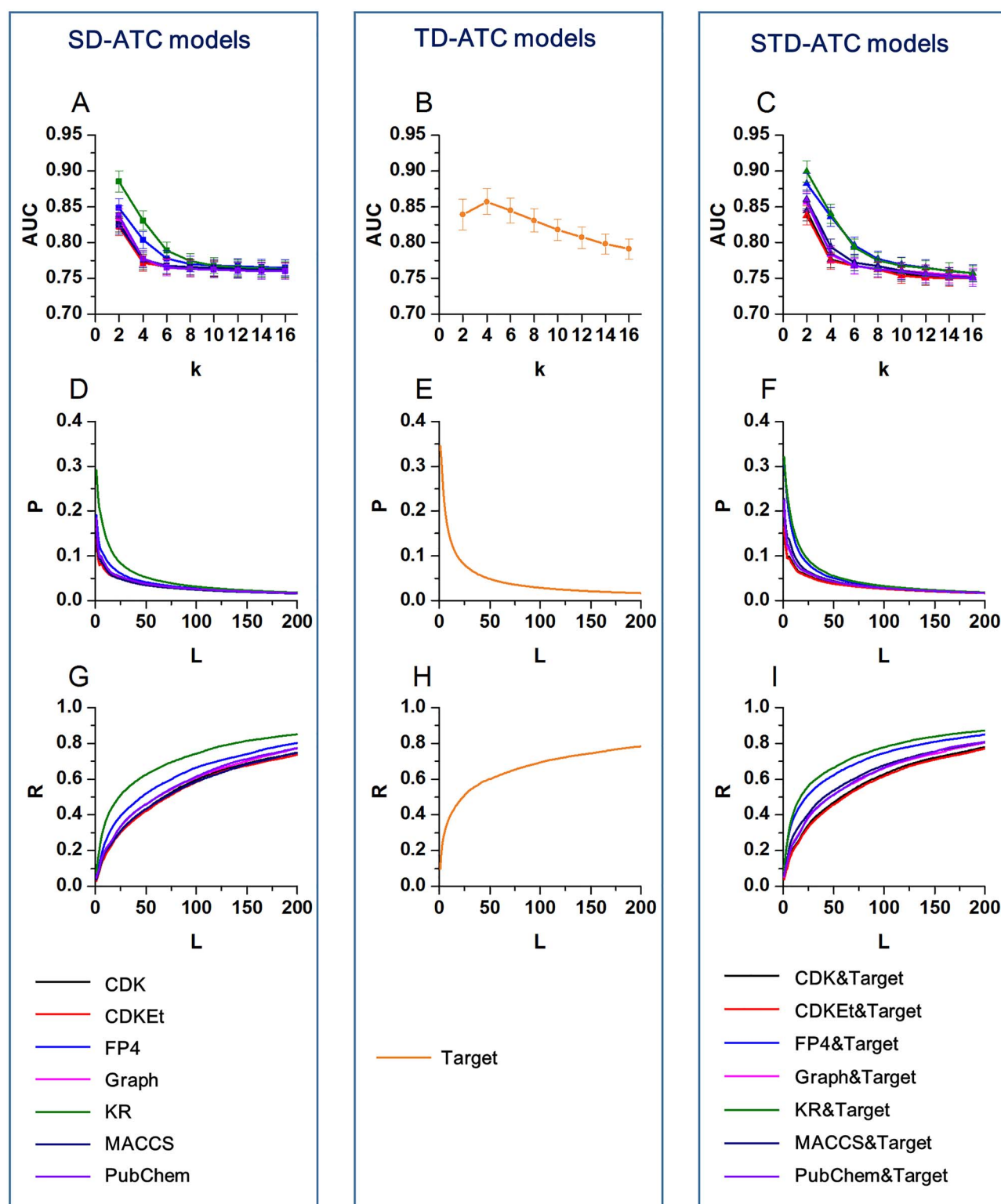


Figure 4. The AUC-k, P-L and R-L curves of different prediction models in 10-fold cross validation. (A) AUC-k curves of SD-ATC. (B) AUC-k curves of TD-ATC. (C) AUC-k curves of STD-ATC. (D) P-L curves of SDA. (E) P-L curves of TD-ATC. (F) P-L curves of STD-ATC. (G) R-L curves of SD-ATC. (H) R-L curves of TD-ATC. (I) R-L curves of STD-ATC.

KR&Target and FP4&Target models were better than the other STD-ATC models, too. The validation indicators of P, R and AUC of different models were listed in Table 3 when $k=2$ and $L=15$. From Figure 5, the same results could be found, that is, KR, FP4, KR&Target and FP4&Target were better models than the others. Most importantly, the STD-ATC model is always better than the corresponding SD-ATC model. The AUC values of CDK&Target, FP4&Target, Graph&Target, KR&Target, MACCS&Target and PubChem&Target models were all higher than the target model. Therefore, when protein target and substructure information were simultaneously used to construct the models and predict the ATC codes, the accuracy of models could be improved.

Evaluation of the model generalization ability

To evaluate the generalization ability of our models, two external data sets were used to assess the performance of different prediction models. From Figures S2A–C and S3A–B (see Supplementary Data available online at <http://bib.oxfordjournals.org/>), we found that $k=2$ was the optimal value, which was consistent with the results of training set. From Figures S2D–I and S3C–F (see Supplementary Data available online at <http://bib.oxfordjournals.org/>), we realized that P decreased and R increased as L increased, which was also conformed to the results of training set. At the same time, the KR and FP4 models were the top two SD-ATC models, while KR&Target and FP4&Target models were the best two STD-ATC models.

When comparing Figure 5B and C, we could see that the AUC value of each model in external set 1 was higher than that of external set 2. That is because of the absence of target information in external set 2, where the resource of those drugs could not be diffused through TD-ATC network. By the same token, as to the comparison of the SD-ATC and STD-ATC models, there were less improvement of STD-ATC models of external set 2 when comparing Figure 5B and C. On the contrary, when focusing on external set 1, it was easy to see that the AUC values of STD-ATC models were better than those of SD-ATC and TD-ATC models. Therefore, we could summarize that target information is vital for ATC code prediction. When compounds possessed experimentally validated protein targets, more comprehensive information should be considered to improve ATC code prediction, including substructures and protein targets of compounds. Eventually, the KR&Target model was the best.

Case study: drug repositioning for heart failure

Computational prediction

From the results of 10-fold validation and two external validations, with KR fingerprint the STD-ATC model was the optimal one. Therefore, the KR&Target model was applied to predict ATC codes for approved Drugs. The final number of 'approved drugs' for ATC code prediction is 2309. According to known HF-related ATC codes and the predicted ATC codes of each drug, Fisher's exact test and Benjamini Hochberg adjustment were conducted, 25 drugs with P-value less than 0.05 were predicted with potential HF therapeutic effects.

Experimental verification

To evaluate the accuracy of our prediction model, 22 of 25 drugs were obtained from our in-house drug library or purchased as supplement for the evaluation of their cardioprotective properties by two cell injury models *in vitro*. All drugs were tested at 10 and 100 μM , respectively (Table S4, see Supplementary Data available online at <http://bib.oxfordjournals.org/>). Here are the results.

Hypoxia model. Hypoxia is an important related factor during the progression of heart failure. We used a single-factor experiment model that causes hypoxia injury [38]. Dexmedetomidine was chosen as a positive control for its good performance in reducing apoptosis and myocardial injury [39]. As shown in Figure 6A, after 24-h hypoxia treatments, 12 drugs significantly reduced LDH release compared with model group, showing potent cardioprotective activity. Especially, DB00871, DB06707, DB09086 and DB11284 exhibited cytoprotection at both 10 μM and 100 μM .

OGD model. Ischemic myocardium leads to myocardial cell damage and ultimately heart failure. Oxygen–glucose deprivation model simulated the environment of myocardial ischemia *in vitro*. Cell viability was measured after OGD treatments. As described in Figure 6B, 13 drugs increased cell viability significantly compared with model group (45.29%) after 3-h OGD treatments. The cell viabilities and cardioprotective activities of the 13 old drugs were shown in Table 4. Among these compounds, DB00968, DB02701, DB06707 and DB11268 showed better cardioprotective activity than nicorandil, an effective drug in acute or chronic ischemic heart disease, congestive heart failure and arrhythmia [40], with cardioprotective activity 21.9914, 35, 10.18 and 13.12%, respectively.

Based on the above experimental results, 8 old drugs among the 22 ones had shown mild to potent cardioprotective activities in both hypoxia model and OGD model, which confirmed our network-based inference models for the prediction of ATC codes.

Discussion

In this study, we proposed a novel network-based inference method to predict the 2nd-, 3rd- and 4th-level ATC codes of drugs. Three types of tripartite networks, namely, SD-ATC, TD-ATC and STD-ATC, were constructed and 15 different models were built.

Our prediction models were validated comprehensively by 10-fold cross validation, external validation set 1 with known protein targets and external validation set 2 without known target information. Simultaneously, different models were compared to each other. After 10-fold cross validation and two external validations, we obtained the best model, STD-ATC with KR as fingerprint, to predict ATC codes for all the approved drugs from DrugBank. Then we used Fisher's exact test for the virtual screening of potential old drugs with cardioprotective activities. Importantly, both the hypoxia and OGD models confirmed the predictive results with 8 of 22 (25 old drugs were predicted, but only 22 old drugs were actually purchased for experimental assays) approved drugs exhibiting cardioprotective effects (Table S4, see Supplementary Data available online at <http://bib.oxfordjournals.org/>). Therefore, combining network-based inference, statistical test and experimental validation, the STD-ATC model with KR as fingerprint showed the abilities not only for the prediction of ATC codes but also for understanding of their pharmacological effects.

Compared with previously published methods for the prediction of ATC codes, such as SuperPred [19], NetPredATC [20] and ATC-NLSP [24], our method exhibited some advantages, though a detailed comparison between them by indicators is very difficult due to diverse data sets used. First of all, with the prediction models, we could explain the relationships between the 2nd-, 3rd- and 4th-level ATC codes and therapeutic and pharmacological effects. By predicting the ATC codes for old drugs, we inferred new imbedded effects for old

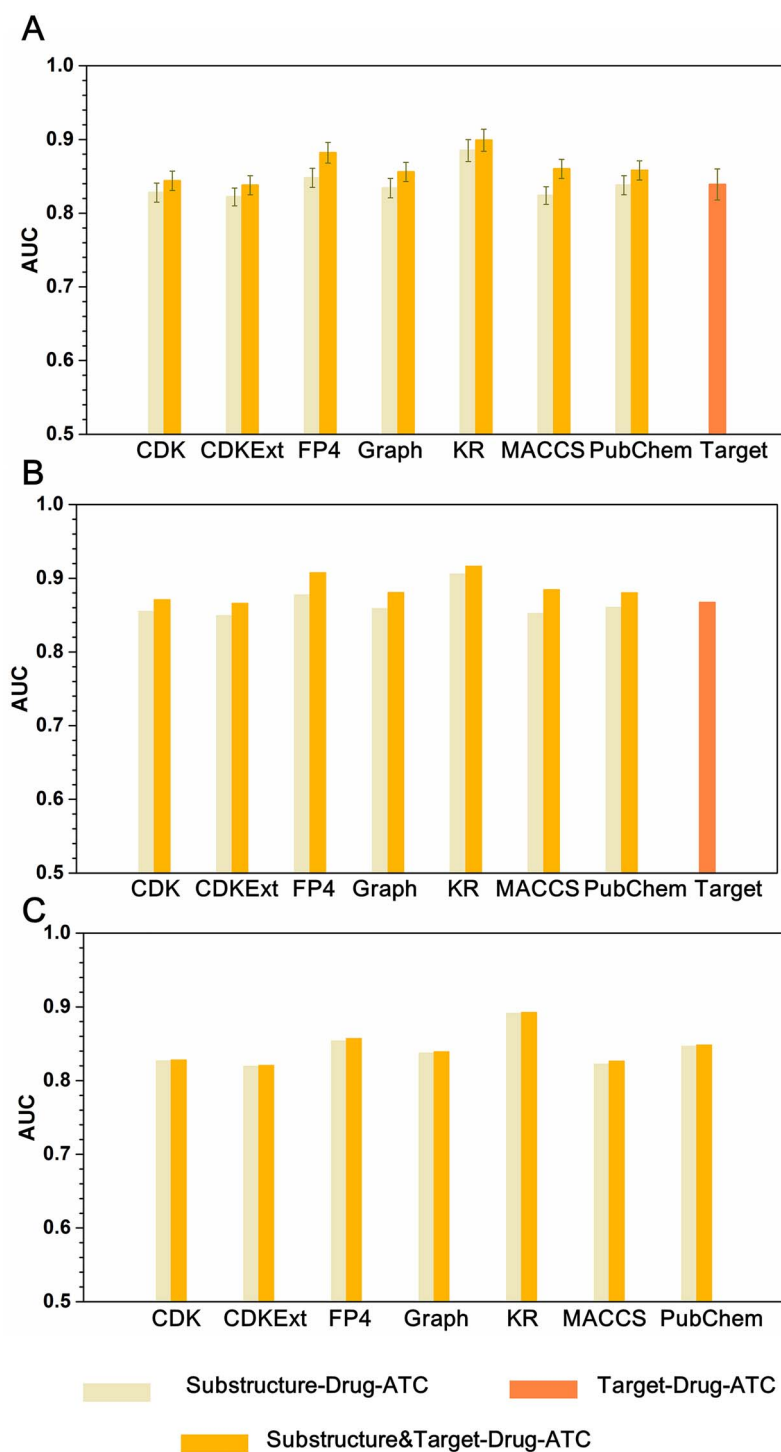


Figure 5. The histogram of the AUC value in SD-ATC, TD-ATC and STD-ATC models. (A) 10-fold cross validation. (B) External validation set 1. (C) External validation set 2.

drugs and confirmed our prediction results by two *in vitro* experimental assays. Therefore, we made the prediction that ATC codes have biological significance. However, the previously published studies did not apply their prediction models in drug discovery and had no experimental validation yet. Especially, Chen *et al.* [23] just predicted the 1st-level ATC code, which is not helpful for drug research. Secondly, we collected a comprehensive and up-to-date training set and two independent external validation sets from DrugBank. The first-hand

data are the most important for computational prediction. However, Wang *et al.* [20] (NetPredATC) and Wang *et al.* [24] (ATC-NLSP) constructed their prediction models based on second-hand data, which were not helpful for building high-quality prediction models. Thirdly, when comparing SD-ATC, TD-ATC and STD-ATC models, we found that STD-ATC model was the best, which was consistent with the findings in Wang's work [20]. Both of us found that combination of the substructure and target information could offer better

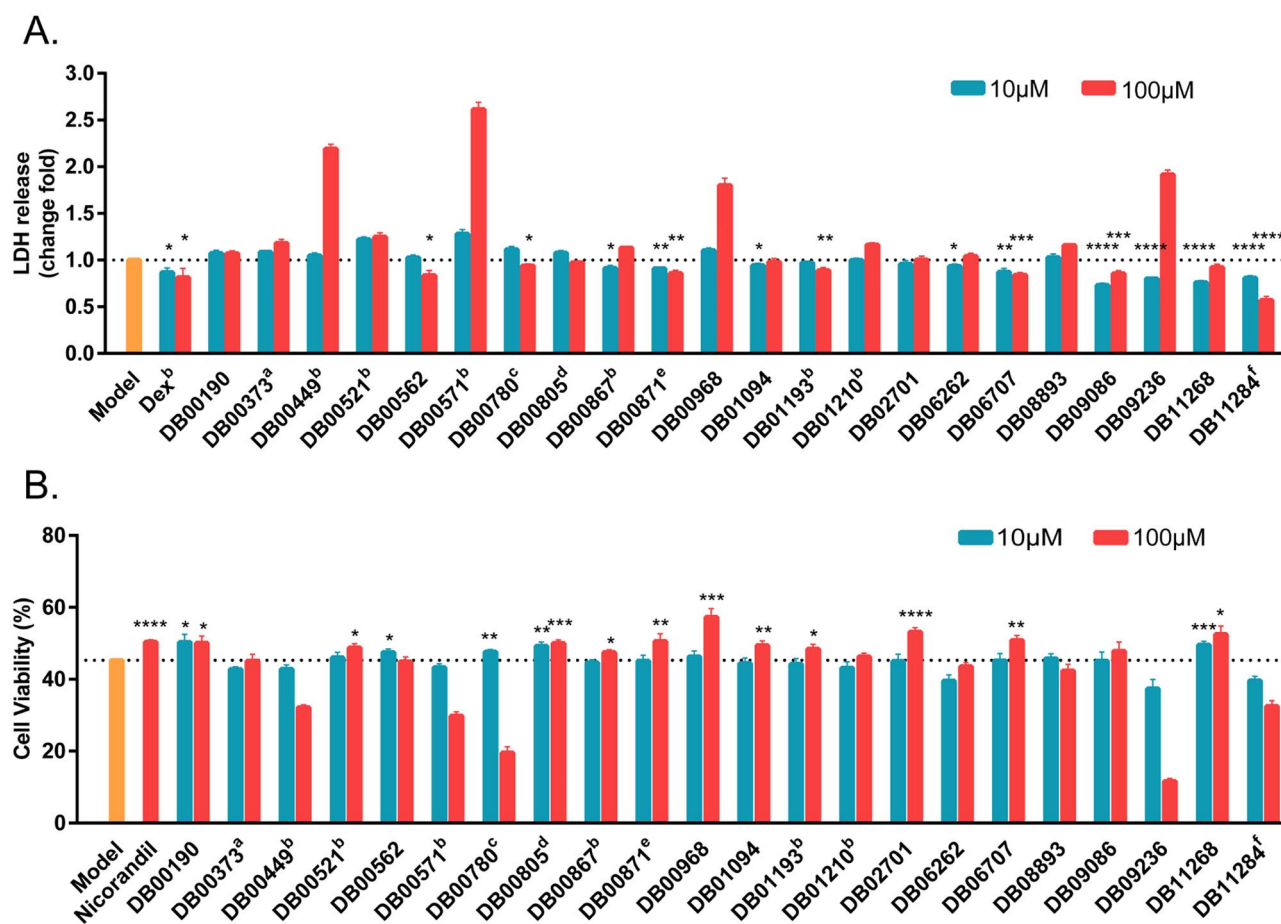


Figure 6. Results of experimental validation. Salt type: ^amaleate, ^bhydrochloride, ^csulfate, ^ddihydrochloride, ^ehemisulfate, ^fsodium salt. (A) The LDH release of H9C2 (2–1) cell after 24-h hypoxia. Dexmedetomidine as positive drug. (B) The cell viability of H9C2 (2–1) after 3-h oxygen-glucose deprivation. Nicorandil as positive drug. Data is shown as the mean \pm SEM of at least three independent experiments: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$ versus model group.

prediction. Among all STD-ATC models, we realized that the model with KR as fingerprint was the best. Though Wang *et al.* adopted chemical and protein similarity profile, they did not evaluate different similarity methods, whereas SuperPred only considered chemical similarity [19]. Comparing with SuperPred of Dunkel *et al.* [19], one shortage of our method is that we did not provide a webserver for wide use yet. We will improve this by developing a webserver for ATC code prediction in the near future.

Our method provided a new approach for drug repositioning by the prediction of chemical ATC code. Comparing with target-based approach for drug repositioning, we could screen the effective old drugs from the therapeutic or pharmacological level firstly, to avoid the possibility that a drug was just effective at molecular level. Our drug repositioning strategy was also better than those based on machine learning methods. As we know, machine learning methods need balanced positive and negative data for model building. However, high-quality negative biological data are often difficult to be obtained for the reason that corporations or scientists tend to report effective data rather than futile results. Nevertheless, our network-based inference method was independent of negative data. By utilizing known drug-ATC code network data proposed by the WHO, the potential drug-ATC code relationships could be predicted.

Our method could also be applied in the prediction of ATC codes for new chemicals, so it has potential applications in

virtual screening (Figure 3B). The ordinary virtual screening procedure could be classified into two types: structure-based and ligand-based. The structure-based approach needs known three-dimensional crystal structures of specific targets, and the connections between targets and diseases should be clear. The ligand-based approach needs known ligand data as to a given target. Therefore, our ATC code prediction method offered an alternative virtual screening approach by predicting the potential ATC codes for chemicals from commercial or in-house data. Our method was not limited with the three-dimensional structures of targets or known ligand data. We also finished a project for natural product research. We know that ingredients of natural products are usually complicated and the certification of the effective ingredients is laborious and intricate. According to the predicted 2nd-, 3rd- and 4th-level ATC codes of all ingredients, we could discern the latent effective ones effortlessly.

Conclusions

ATC code contains pharmacological and therapeutic information of drugs; hence the prediction of ATC code could provide key information for drug repositioning. In this study, we developed *in silico* models for the prediction of ATC codes via network-based inference approaches. Among the three types

of models, STD-ATC models outperformed the SD-ATC and TD-ATC ones after 10-fold cross validation and two external validations. Especially, among the 15 models, using KR as fingerprint, the STD-ATC model obtained the best performance with AUC of 0.899 ± 0.015 , 0.916 and 0.893 in 10-fold cross validation, external validation set 1 and external validation set 2, respectively.

Then, to further validate the STD-ATC model with KR fingerprint, we conducted a case study to predict ATC codes among approved drugs for potential effects on heart failures. We employed both hypoxia and OGD *in vitro* models to validate the computational results. Ultimately, 8 of the 22 approved drugs exhibited cardioprotective effects on both hypoxia and OGD models (Table S4, see [Supplementary Data](#) available online at <http://bib.oxfordjournals.org/>). These results demonstrated that our models for the prediction of ATC codes would be very helpful for drug repositioning and phenotype-based virtual screening.

Key Points

- Prediction of the 2nd-, 3rd- and 4th-level ATC codes is a novel strategy for drug repositioning.
- Three types of models were built and compared comprehensively, among which STD-ATC model outperformed SD-ATC and TD-ATC ones.
- With KR as fingerprint, the STD-ATC model was used to predict ATC codes of 25 old drugs for potential applications in the treatment of heart failure.
- Both hypoxia and OGD models experimentally confirmed that 8 of the predicted 22 old drugs exhibited cardioprotective effects.

Supplementary data

Supplementary data are available online at <http://bib.oxfordjournals.org/>.

Funding

National Key Research and Development Program of China (2016YFA0502304); the National Natural Science Foundation of China (81673356, 81872800); and the Chinese Special Fund for State Key Laboratory of Bioreactor Engineering (2060204).

Conflict of interest

All the authors have no conflict of interest.

References

1. Dowden H, Munro J. Trends in clinical success rates and therapeutic focus. *Nat Rev Drug Discov* 2019;**18**:495–6.
2. Ashburn TT, Thor KB. Drug repositioning: identifying and developing new uses for existing drugs. *Nat Rev Drug Discov* 2004;**3**:673–83.
3. Govindaraj RG, Naderi M, Singha M, et al. Large-scale computational drug repositioning to find treatments for rare diseases. *NPJ Syst Biol Appl* 2018;**4**:13.
4. Pushpakom S, Iorio F, Eyers PA, et al. Drug repurposing: progress, challenges and recommendations. *Nat Rev Drug Discov* 2019;**18**:41–58.

5. WHO Expert Committee on the Selection and Use of Essential Medicines, World Health Organization. The selection and use of essential medicines. In: *Report of the WHO expert committee, 2005 (including the 14th model list of essential medicines)*, 2006.
6. Olson T, Singh R. Predicting anatomic therapeutic chemical classification codes using tiered learning. *BMC Bioinformatics* 2017;**18**:266.
7. Wishart DS, Feunang YD, Guo AC, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res* 2018;**46**:D1074–82.
8. Kim S, Thiessen PA, Bolton EE, et al. PubChem substance and compound databases. *Nucleic Acids Res* 2016;**44**:D1202–13.
9. Bailey CJ. Metformin: historical overview. *Diabetologia* 2017;**60**:1566–76.
10. Montinari MR, Minelli S, De Caterina R. The first 3500 years of aspirin history from its roots—a concise summary. *Vascul Pharmacol* 2019;**113**:1–8.
11. Garcia EY. Flumamine, a new synthetic analgesic and anti-flu drug. *J Philipp Med Assoc* 1950;**26**:287–93.
12. Holman RR, Paul SK, Bethel MA, et al. 10-year follow-up of intensive glucose control in type 2 diabetes. *N Engl J Med* 2008;**359**:1577–89.
13. Quick A. Bleeding-time after aspirin ingestion. *The Lancet* 1968;**291**:50.
14. Algra AM, Rothwell PM. Effects of regular aspirin on long-term cancer incidence and metastasis: a systematic comparison of evidence from observational studies versus randomised trials. *Lancet Oncol* 2012;**13**:518–27.
15. Hopkins AL. Network pharmacology: the next paradigm in drug discovery. *Nat Chem Biol* 2008;**4**:682.
16. Cheng F, Liu C, Jiang J, et al. Prediction of drug-target interactions and drug repositioning via network-based inference. *PLoS Comput Biol* 2012;**8**:e1002503.
17. Wu Z, Cheng F, Li J, et al. SDTNBI: an integrated network and chemoinformatics tool for systematic prediction of drug-target interactions and drug repositioning. *Brief Bioinform* 2017;**18**:333–47.
18. Liu Z, Guo F, Gu J, et al. Similarity-based prediction for anatomical therapeutic chemical classification of drugs by integrating multiple data sources. *Bioinformatics* 2015;**31**:1788–95.
19. Dunkel M, Gunther S, Ahmed J, et al. SuperPred: drug classification and target prediction. *Nucleic Acids Res* 2008;**36**:W55–9.
20. Wang YC, Chen SL, Deng NY, et al. Network predicting drug's anatomical therapeutic chemical code. *Bioinformatics* 2013;**29**:1317–24.
21. Yamanishi Y, Araki M, Gutteridge A, et al. Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics* 2008;**24**:i232–40.
22. Chen FS, Jiang ZR. Prediction of drug's anatomical therapeutic chemical (ATC) code by integrating drug-domain network. *J Biomed Inform* 2015;**58**:80–8.
23. Chen L, Liu T, Zhao X. Inferring anatomical therapeutic chemical (ATC) class of drugs using shortest path and random walk with restart algorithms. *Biochim Biophys Acta Mol Basis Dis* 1864;**2018**:2228–40.
24. Wang X, Wang Y, Xu Z, et al. ATC-NLSP: prediction of the classes of anatomical therapeutic chemicals using a network-based label space partition method. *Front Pharmacol* 2019;**10**:971.

25. O'Boyle NM, Banck M, James CA, et al. Open babel: an open chemical toolbox. *J Chem* 2011;**3**:33.
26. Yap CW. PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. *J Comput Chem* 2011;**32**:1466–74.
27. Bento AP, Gaulton A, Hersey A, et al. The ChEMBL bioactivity database: an update. *Nucleic Acids Res* 2014;**42**:D1083–90.
28. Gilson MK, Liu T, Baitaluk M, et al. BindingDB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Res* 2016;**44**:D1045–53.
29. Harding SD, Sharman JL, Faccenda E, et al. The IUPHAR/BPS guide to PHARMACOLOGY in 2018: updates and expansion to encompass the new guide to IMMUNOPHARMACOLOGY. *Nucleic Acids Res* 2018;**46**:D1091–106.
30. Roth BL, Lopez E, Patel S, et al. The multiplicity of serotonin receptors: uselessly diverse molecules or an embarrassment of riches? *Neuroscientist* 2000;**6**:252–62.
31. Fliri AF, Loging WT, Thadeio PF, et al. Analysis of drug-induced effect patterns to link structure and side effects of medicines. *Nat Chem Biol* 2005;**1**:389–97.
32. Wu Z, Lu W, Wu D, et al. In silico prediction of chemical mechanism of action via an improved network-based inference method. *Br J Pharmacol* 2016;**173**:3372–85.
33. Cheng F, Zhou Y, Li W, et al. Prediction of chemical-protein interactions network with weighted network-based inference method. *PLoS One* 2012;**7**:e41064.
34. Wu Z, Li W, Liu G, et al. Network-based methods for prediction of drug-target interactions. *Front Pharmacol* 2018;**9**:1134.
35. Fang J, Wu Z, Cai C, et al. Quantitative and systems pharmacology. 1. In silico prediction of drug-target interactions of natural products enables new targeted cancer therapy. *J Chem Inf Model* 2017;**57**:2657–71.
36. Rossignol P, Hernandez AF, Solomon SD, et al. Heart failure drug treatment. *The Lancet* 2019;**393**:1034–44.
37. Zannad F. Rising incidence of heart failure demands action. *The Lancet* 2018;**391**:518–9.
38. Du J, Liu J, Zhen J, et al. Astragaloside IV protects cardiomyocytes from hypoxia-induced injury by down-regulation of lncRNA GAS5. *Biomed Pharmacother* 2019;**116**:109028.
39. Wu Z, Chu S, Gui P, et al. Effects of dexmedetomidine pretreatment on heart function and cell apoptosis after myocardial ischemia reperfusion injury in rats. *J Clin Anesth* 2015;**31**:901–4.
40. Horinaka S. Use of nicorandil in cardiovascular disease and its optimization. *Drugs* 2011;**71**:1105–19.