OPEN ACCESS Freely available online

PLoS one

Prediction of Chemical-Protein Interactions Network with Weighted Network-Based Inference Method

Feixiong Cheng, Yadi Zhou, Weihua Li, Guixia Liu, Yun Tang*

Shanghai Key Laboratory of New Drug Design, School of Pharmacy, East China University of Science and Technology, Shanghai, China

Abstract

Chemical-protein interaction (CPI) is the central topic of target identification and drug discovery. However, large scale determination of CPI is a big challenge for *in vitro* or *in vivo* experiments, while *in silico* prediction shows great advantages due to low cost and high accuracy. On the basis of our previous drug-target interaction prediction *via* network-based inference (NBI) method, we further developed node- and edge-weighted NBI methods for CPI prediction here. Two comprehensive CPI bipartite networks extracted from ChEMBL database were used to evaluate the methods, one containing 17,111 CPI pairs between 4,741 compounds and 97 G protein-coupled receptors, the other including 13,648 CPI pairs between 2,827 compounds and 206 kinases. The range of the area under receiver operating characteristic curves was 0.73 to 0.83 for the external validation sets, which confirmed the reliability of the prediction. The weak-interaction hypothesis in CPI network was identified by the edge-weighted NBI method. Moreover, to validate the methods, several candidate targets were predicted for five approved drugs, namely imatinib, dasatinib, sertindole, olanzapine and ziprasidone. The molecular hypotheses and experimental evidence for these predictions were further provided. These results confirmed that our methods have potential values in understanding molecular basis of drug polypharmacology and would be helpful for drug repositioning.

Citation: Cheng F, Zhou Y, Li W, Liu G, Tang Y (2012) Prediction of Chemical-Protein Interactions Network with Weighted Network-Based Inference Method. PLoS ONE 7(7): e41064. doi:10.1371/journal.pone.0041064

Editor: Eugene A. Permyakov, Russian Academy of Sciences, Institute for Biological Instrumentation, Russian Federation

Received April 20, 2012; **Accepted** June 16, 2012; **Published** July 16, 2012

Funding: This work was supported by the 863 Project (Grant 2012AA020308), the National Natural Science Foundation of China (Grant 21072059), the 111 Project (Grant B07023), the Fundamental Research Funds for the Central Universities (WY1113007), and the Shanghai Committee of Science and Technology (11DZ2260600). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: ytang234@ecust.edu.cn

Introduction

Over the past decade, the productivity of drug research and development (R&D) seems to be decreasing [1]. Richard *et al.* stated that at present more than 800 drugs are in clinical development for cancer indications and the current success rate in bringing drugs to the markets remains only in the range of 5–8% [2]. One reason about R&D decrease might be due to the domination of ''one-disease-one-drug-one-target' paradigm [1]. Several clinical investigations confirmed that most drugs act on multiple targets rather than one target, that is, drug polypharmacology [1,3]. For example, tamoxifen, which is an approved drug used to treat breast cancer for more than 30 years, has been found to be effective in experimental models of cutaneous and visceral leishmaniasis [4]. Several well-known drugs such as thalidomide, sildenafil, bupropion and fluoxetine were found new uses beyond their original approved therapeutic indications [5].

Study of chemical-protein interactions (CPI) network is an important topic toward elucidation of protein functions, understanding of molecular mechanisms inside the cell and drug repositioning. It is both time-consuming and costly to identify CPI by experiments alone. As a complement, *in silico* method could provide us with very useful information in a predictable, reliable, less costly and timely manner. Various *in silico* methods have been proposed to address the CPI prediction. The classical methods can be classified into ligand-based and target-based ones. For example, Humberto *et al.* developed a multi-target QSAR classifier and built a web server for CPI prediction [6]. Another widely used target-based method is reverse molecular docking. Several web servers, such as TarFisDock [7], DRAR-CPI [8] for drug discovery and CPI prediction have been developed. However, this method cannot be applied to targets whose three-dimensional (3D) structures are unresolved, especially for membrane proteins like G protein-coupled receptors (GPCRs), and were limited usage due to time-costly and the inaccuracy of the scoring functions.

Recently, several new methods, such as computational chemogenomics, phenotype-based and network-based diffusion methods were successful proposed for CPI prediction [9–23]. Yamanishi *et al.* developed a bipartite graph learning method for drug-target interaction (DTI) prediction [9]. Recently, Yamanishi *et al.* further developed DTI prediction method by integrating chemical, genomic and pharmacological spaces [21]. Though high overall predictive accuracy was yielded in the Yamanishi's work, the sensitivity was anomaly low and the method was not validated experimentally. Wang *et al.* developed a computational chemogenomics method from protein primary sequences and used it to identify several new ligands for four targets (i.e., GPR40, SIRT1, p38, and GSK-3β) validated by experimental assays [11]. The drawback of chemogenomics method is that there are a huge number of samples to be classified, which increase the computational complexity. Another bottleneck is the lack of the benchmark negative CPI pairs and it easily results in high false positive rate. Our recent work found that there is high false positive rate in computational chemogenmics method, and the performance of

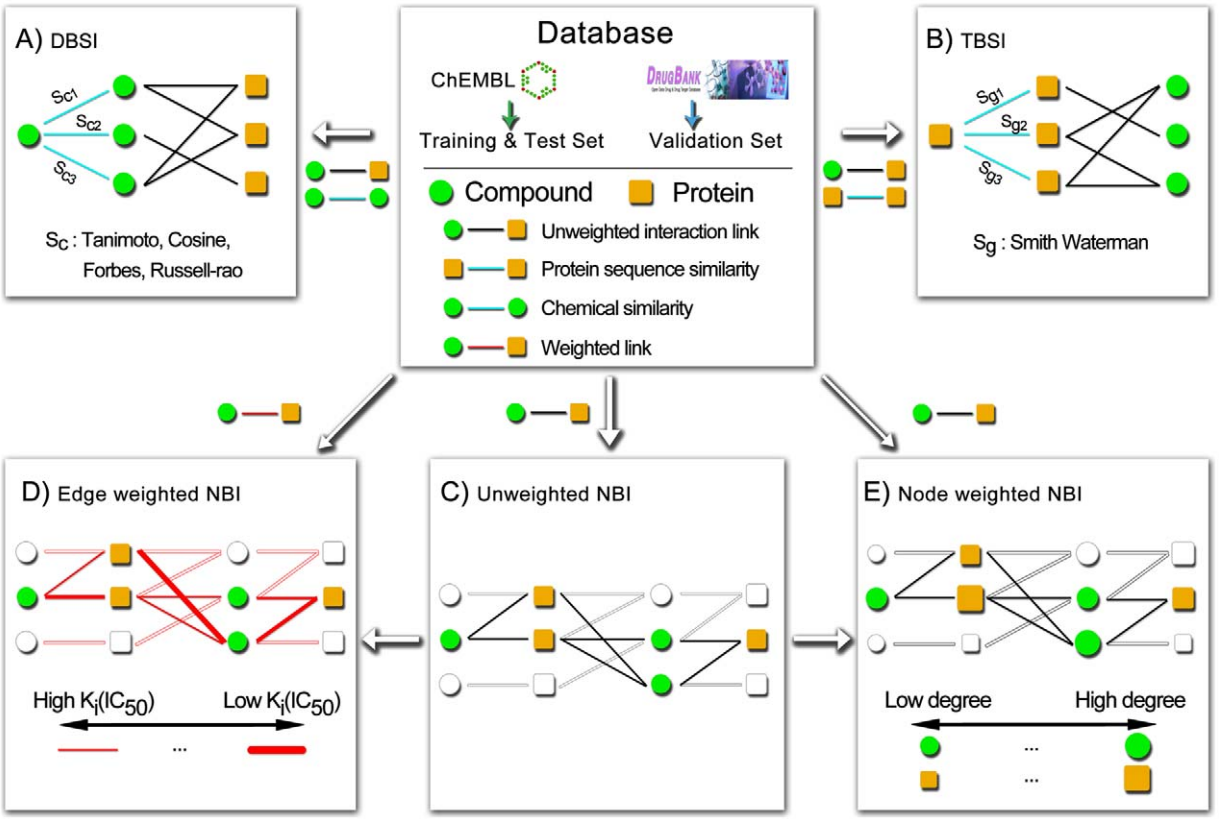

Figure 1. Schematic diagram of our proposed method. (**A**) The drug-based similarity inference (DBSI), (**B**) the target-based similarity inference (TBSI) and (**C**) the unweighted network-based inference (NBI), (**D**) the edge-weighted NBI (EWNBI) and (**E**) the node-weighted NBI (NWNBI). Green circle: chemical node, gold square: protein node, black line: unweighted interaction link, cyan line: chemical-chemical two-dimensional structural similarity (S_c) or protein-protein Smith Waterman genomic similarity (S_g), red line: weighted edges (thick red line denotes the strong edge with high potency and thin red line denotes the weak edge with low potency).
doi:10.1371/journal.pone.0041064.g001

chemogenmics method was influenced by data set bias and features selection methods [23].

In our previous paper, we reported three supervised inference methods: drug-based similarity inference (DBSI), target-based similarity inference (TBSI) and network-based inference (NBI) methods for DTI prediction and drug repositioning derived from complex network theory [24,25]. With the methods, five known drugs were predicted and experimentally validated to have novel indications on estrogen receptors and dipeptidyl peptidase-IV [26]. However, the methods are only suitable for drugs having known links to targets in the training set and the unweighted DTI network among drug and target nodes was used. Whether the weighted DTI network could improve the predictive accuracy or the method could be extended to general CPI prediction has not been investigated yet.

In this paper, the above-mentioned three methods were further improved. Two new methods, namely node-weighted network-based inference (NWNBI) and edge-weighted network-based inference (EWNBI) were further presented, and four similarity metrics (Tanimoto, Cosine, Forbes and Russell-rao) were explored in the DBSI method systematically, for CPI prediction and drug repositioning. The methods were then examined with two comprehensive CPI databases targeting GPCRs and kinases, respectively. The new targets were further predicted for five example known drugs, and experimental evidences to support the predictions were provided.

Materials and Methods

Data Preparation

Two comprehensive CPI data sets were collected from the ChEMBL database (https://www.ebi.ac.uk/chembl/, accessed in May. 2010) [27]. The initial database includes 1,195,368 compounds and more than 8,000 targets from various species. Here, we only focused on two pharmacologically important families, GPCRs and kinases. The data sets were refined with the following criteria: (1) only human test data were selected; (2) only those with K_i or IC_{50} values less than 10 μM were extracted; (3) proteins connected with less than three active compounds were excluded; (4) proteins with non-standard amino acids, DNA, RNA, or sequence length less than 100 residues were removed; and (5) nonorganic chemicals and chemicals with molecular weight less than 100 Dalton or more than 600 Dalton were also excluded. All compounds in SMILES format and proteins sequence in FASTA format were extracted from ChEMBL.

Network Construction

The methods adopted in this paper are to prioritize unconnected candidate proteins for a given chemical, or prioritize unconnected candidate chemicals for a given protein, which derived from the recommendation algorithms of complex network theory [26]. We constructed two comprehensive CPI bipartite networks (or graphs) to represent the data in chemical nodes, protein nodes and their physical interactions. Denoting the

chemical set as $C=\{c_1,c_2,\ldots,c_n\}$and the protein set as $P=\{p_1,p_2,\ldots,p_m\}$, the CPI binary pairs can be described as a bipartite CPI graph $G(C,P,E)$, where $E=\{e_{ij}: c_i\in C,p_j\in P\}$. A link is drawn between c_i and p_j only if the K_i or IC_{50} was less than 10 μM between c_i and p_j. The CPI bipartite graph can be presented by an $n\times m$ adjacent matrix$\{a_{ij}\}$, where $a_{ij}=1$ when K_i or IC_{50} value less than 10 μM, otherwise$a_{ij}=0$.

Methods Development

In our previous work, we proposed three inference methods, i.e. DBSI, TBSI and NBI, to predict DTI. In this study, we managed to improve the NBI method with weighted nodes or edges. The entire workflow was illustrated in Figure 1.

Drug-Based Similarity Inference (DBSI). The DBSI method was designed based on the hypothesis that two chemicals with similar chemical structures may exhibit similar bioactivities (Figure 1A), which was described in our previous work [26]. For a CPI pair c_i-p_j, if c_i has not interacted with p_j yet, the predicted score by this method is given as:

$$v_{ij}^{D}=\frac{\sum_{l=1,l\neq i}^{n} S_c(c_i,c_l)a_{lj}}{\sum_{l=1,l\neq i}^{n} S_c(c_i,c_l)} \tag{1}$$

$S_c(c_i,c_l)$ indicates two-dimensional (2D) chemical structural similarity between chemicals c_i and c_l. In this study, four different chemical structure similarity metrics, namely Tanimoto, Cosine, Forbes and Russell-rao were systemically evaluated using MACCS keys, freely available from OpenBabel (version 2.3.0) [28]. The further descriptions about four similarity metrics were given in the work of Willett *et al.* [29].

Target-Based Similarity Inference (TBSI). The TBSI method was designed based on the hypothesis that two proteins with similar genomic space may exhibit similar biology function (Figure 1B). For any CPI pair c_i-p_j, if c_i does not connect with p_j in the bipartite graph, the predicted score by this method is given as:

$$v_{ij}^{T}=\frac{\sum_{l=1,l\neq j}^{m} S_g(p_j,p_l)a_{il}}{\sum_{l=1,l\neq j}^{m} S_g(p_j,p_l)} \tag{2}$$

$S_g(p_j,p_l)$ indicates the genomic sequence similarity between two proteins p_j and p_l. The sequence similarity between protein p_j and p_l was computed by the Smith-Waterman scores [30].

Unweighted Network-based Inference (NBI). Considering the bipartite graph $G(C,P,E)$, we applied a mass diffusion-based method to obtain the predicted list. For a given chemical c_i, supposing that a kind of resource is initially located in the proteins which are interacted with c_i, the resource will diffuse to all the proteins in the network after the network-based resource allocation process [24,25]. Each protein node averagely distributes its resource to all neighboring chemical nodes and then each chemical redistribute the received resource to all neighboring protein nodes. The finial resource on the proteins that are not connected with the chemical c_i in G could be considered as the score of each protein, and the proteins with high score are more likely to interact with c_i. Figure 1C gives a simple example to illustrate the network-based resource allocation process. It shows the initial resource of a_{ij} between c_i (green cycle) and p_j (orange square) followed as:

$$a_{ij}=\begin{cases}1 & K_i(IC_{50})\leq 10\mu M\\ 0 & K_i(IC_{50})>10\mu M\end{cases} \tag{3}$$

Denoting $F_{0n\times m}$ as the initial resource matrix (adjacency matrix) and $F_{0ij}=a_{ij}$, $R_{n\times n}$ as the total resource (degree) of each chemical and $R=diag\left(\sum_{j=1}^{m}a_{1j},\ \sum_{j=1}^{m}a_{2j},\ \ldots,\ \sum_{j=1}^{m}a_{nj}\right)$, $H_{m\times m}$ as the total resource (degree) of each protein and $H=diag\left(\sum_{i=1}^{n}a_{i1},\ \sum_{i=1}^{n}a_{i2},\ \ldots,\ \sum_{i=1}^{n}a_{im}\right)$, the final resource matrix will be obtained as $F_{1n\times m}$, and $F_1=F_0W_{m\times m}$ or $F_1^T=F_0^TW_{n\times n}$, where transfer matrix $W_{m\times m}=(F_0H^{-1})^T(R^{-1}F_0)$ or $W_{n\times n}=(R^{-1}F_0)(F_0H^{-1})^T$.

Edge Weighted Network-based Inference (EWNBI). In the above unweighted NBI method, we only consider the binary CPI pairs among nodes. However, the edges among chemicals and proteins are naturally weighted in the real biology world. For the EWNBI method, each edge of CPI network was weighted by the potency ($x_{ij}=-\log_{10}(K_i$ (or $IC_{50})/100$ μM)) of binding affinity (K_i) or inhibitory activity (IC_{50}) of the physical interactions between the chemical node c_i and protein node p_j.

Figure 1D gives a simple example to illustrate the edges weighted network-based resource allocation process. The initial resource of a'_{ij} between c_i (green cycle) and p_j (orange square) were defined as follows:

$$a'_{ij}=\begin{cases}x_{ij} & K_i(IC_{50})\leq 10\mu M\\ 0 & K_i(IC_{50})>10\mu M\end{cases} \tag{4}$$

Denoting $F'_{0n\times m}$ as the initial resource matrix and $F'_{0ij}=a_{ij}$, $R'_{n\times n}$ as the total resource of each chemical and $R'=diag\left(\sum_{j=1}^{m}a'_{1j},\ \sum_{j=1}^{m}a'_{2j},\ \ldots,\ \sum_{j=1}^{m}a'_{nj}\right)$, $H'_{m\times m}$ as the total resource of each protein and $H'=diag\left(\sum_{i=1}^{n}a'_{i1},\ \sum_{i=1}^{n}a'_{i2},\ \ldots,\ \sum_{i=1}^{n}a'_{im}\right)$, the final resource matrix will be obtained as $F'_{1n\times m}$, and $F'_1=F'_0W'_{m\times m}$ or $F'^T_1=F'^T_0W'_{n\times n}$, where transfer matrix $W'_{m\times m}=(F'_0H'^{-1})^T(R'^{-1}F'_0)$ or $W'_{n\times n}=(R'^{-1}F'_0)(F'_0H'^{-1})^T$.

Node Weighted Network-based Inference (NWNBI). Compared to the earlier unweighted NBI method, we use a new expression of initial resource distribution of nodes and take into account the influence of resources associated with the receiver nodes in CPI bipartite network proposed by Jia *et al.* [31]. This method is based on the general knowledge that the hub node with more resources is more difficult to be influenced. Figure 1E illustrates the NWNBI method. For the initial resource matrix, the resources of each chemical and protein node are the same to the unweighted NBI method. The final resource matrix were calculated as$F''_{1n\times m}$, and $F''_{1C}=F_0W''_{m\times m}$ for chemicals and $F''_{1p}=F_0^TW''_{n\times n}$ for proteins, where transfer matrix $W''_{m\times m}=(F_0H^{-1})^T(R^{-1}F_0H''^{-1})$, where $H''=diag\left((\sum_{i=1}^{n}a_{i1})^{\beta},\ (\sum_{i=1}^{n}a_{i2})^{\beta},\ \ldots,\ (\sum_{i=1}^{n}a_{im})^{\beta}\right)$ for chemical or $W''_{n\times n}=(R^{-1}F_0)(R''-1F_0H^{-1})^T$, where $R''=diag\left((\sum_{j=1}^{m}a_{1j})^{\beta},\ (\sum_{j=1}^{m}a_{2j})^{\beta},\ \ldots,\ (\sum_{j=1}^{m}a_{nj})^{\beta}\right)$ for protein, β is a tunable parameter which was used to control the influence. Compared with uniform case, $\beta=0$, a positive β value strengthens the influence of hub nodes, while a negative β value

weakens the influence of hub nodes. The detailed description can be found in Jia's work [31].

Performance Assessment

All performance was assessed based on 10-fold cross validation techniques. In 10-fold cross validation, the entire compound-protein pairs were equally divided into ten cross splits. In each step of cross validation, the model was trained on a set of nine cross validation splits together. The tenth sub-sample set was used as an internal validation set (test set). In order to eliminate the error caused by dividing the data set, all the results were obtained by independent simulation 10 times test. With the randomly splitting, some proteins (or chemicals) maybe just in the test set and the corresponding links couldn't be predicted with our methods, because of no links for these proteins or chemicals in the training set. Such links were not considered in the performance assessment. Mathematically speaking, all methods provide each given chemical with an candidate queue of all its unconnected proteins (C_i (P_a, P_b, …P_m)) or provide each given protein with an candidate queue of all its unconnected chemicals (P_j (C_a, C_b, …C_n)). For each predicted list, we consider the topside links as the most possible candidate CPI. The CPI pairs that were predicted correctly are termed true positive, and the predicted interactions that are not in the test set are referred to as false positive. The area under the receiver operating characteristic curve (AUC) and recall (R) were calculated to assess the performance [26]. In addition, we also calculated the recall enhancement (ER) metric [24].

Results

Network Topology Analysis

Based on the criteria described above, 17,111 CPI among 4,741 unique compounds and 97 GPCRs with K_i less than 10 μM, and 13,648 CPI among 2,827 unique compounds and 206 kinases with IC_{50} less than 10 μM were collected (Table 1). The K_i values, IC_{50} values, SMILES, FASTA, compound ID of all compounds and targeted proteins were given in Tables S1 and S2. The CPI networks of GPCRs and kinases were constructed using a bipartite graph. Figure S1 gives the degree distributions of each chemical and protein. The degree of protein node is the number of chemicals that the protein links with. The degree of chemical node is the number of proteins that the chemical links with. The numbers of chemicals and proteins, the average degree and the sparsity values were given in Table 1. Sparsity is the proportion of the known CPI over all the possible interactions. There is a manifest ligand polypharmacology in GPCRs and kinases (Figure S1).

Relationships among Several Similarities

The heat maps of the compound-compound 2D structural similarity (Tanimoto-scores) and protein sequence similarity (Smith-Waterman scores) were given in Figure S2. The mean values of 2D Tanimoto-scores were 0.458 and 0.444 for ligands of GPCR and kinase, respectively. The range of Smith-Waterman scores of GPCRs was from 0.423 to 0.914 with a mean of 0.517. The range of Smith-Waterman scores of kinases was from 0.383 to 0.996 with a mean of 0.517. From Figure S2, there was diverse ligand chemical space and target coverage of GPCRs and kinases.

We also calculated compound structural activity-relationships (SAR) similarity scores $S_s(c_i,c_j)$ and protein SAR scores $S_s(p_i,p_j)=n_{i,j}/(n_i+n_j-n_{ij})$ using the Tanimoto-score metric, where $n_{i,j}$ is the number of proteins or chemicals that interact with both $c_i(\text{or}p_i)$ and $c_j(\text{or}p_j)$, n_i is the degree of $c_i(\text{or}p_i)$, and n_j is the degree of $c_j(\text{or}p_j)$. The further description about SAR similarity scores was given in Bamborough *et al.* [32]. Figure 2 gave the distribution of compound 2D structural similarities and protein domain sequence similarities against compound and protein SAR similarities. From the box plots, we found two useful features. First, when the compound-compound pairs have the higher Tanimoto scores, they have the higher compound SAR similarities. This feature confirmed the common hypothesis that two ligands with similar structure have the similar biological spectrum [33]. Second, when the protein-protein pairs have the higher Smith-Waterman scores, they have the higher protein SAR similarities. The second feature confirmed the common hypothesis that two proteins with similar structural, functional or evolutionary features will have the similar biological function and bind with similar ligands [34].

Table 1. Statistics of all known chemical-protein interaction pairs of the training set and validation set used in this study.

Data Sets	Targets	N_c	N_p	N_i	N_{dc}	N_{dp}	Sparsity (%)
Training Set	GPCRs	4,741	97	17,111	3.61	176.4	3.72
	Kinases	2,827	206	13,648	4.83	66.3	2.34
Validation Set	GPCRs	92	46	271	2.95	5.89	6.40
	Kinases	188	28	202	1.07	7.21	3.84

N_c: The number of compounds, N_p: The number of proteins, N_i: The number of chemical-protein interactions, N_{dc}: The average degree of compound nodes, N_{dp}: The average degree of protein nodes.
doi:10.1371/journal.pone.0041064.t001

Performance of the Proposed Methods

Unweighted Network-based Inference (NBI). The unweighted NBI method only considers the binary CPI pairs among chemical and protein nodes (Figure 1C). The performance for the test set by 10 simulation times test was summarized in Table 2. The high AUC value of 0.981 ± 0.001 and 0.976 ± 0.002 were yielded for test sets of GPCRs and kinases, respectively. The performance of NBI method was significantly higher than that of the DBSI and TBSI methods, which is in agreement with our previous work [26]. In this study, four similarity metrics, namely Tanimoto, Cosine, Forbes and Russell-rao were systematically investigated. The overall performance of Tanimoto was marginally higher than that of Cosine, Forbes and Russell-rao. Comparing the DBSI and TBSI method (Table 2), the performance of the TBSI method was better than that of the DBSI method, when prioritizing new candidate proteins to a given chemical. The performance of the DBSI method was better than that of the TBSI method, when prioritizing new candidate chemicals to a given protein.

Node Weighted Network-based Inference (NWNBI). As shown in Table 2, the performance of NWNBI was marginally higher than that of the unweighted NBI method. For example, the R value of 0.974 using NWNBI method was marginally higher than 0.969 using unweighted NBI method evaluated on top 5 predicted lists. Figure 3 showed that the performance of test set by simulation 10 times reaches its maximum value at about $\beta=0.3$ for both GPCRs and kinases. Compared with the uniform case of $\beta=0$, a positive β strengthens the influence of hub nodes of chemical or protein, while a negative β weakens the influence of hub nodes. The results indicated that an appropriate increase of

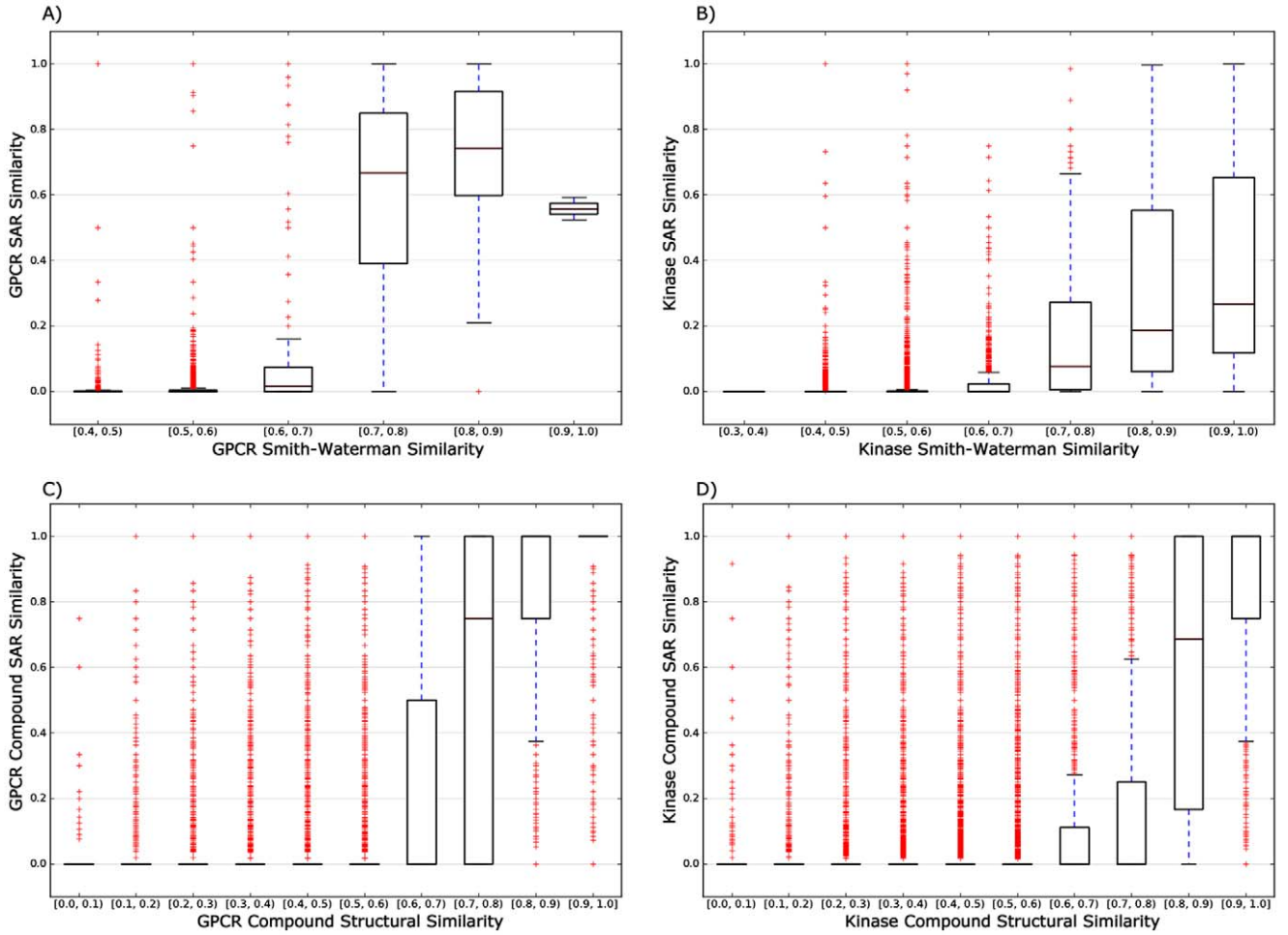

Figure 2. Box plots of compound-compound and protein-protein similarities against compound or protein structure activity-relationship (SAR) similarities. (**A**) protein-protein (GPCRs) sequence similarity (Smith-Waterman scores) against GPCRs SAR similarity, (**B**) protein-protein (kinases) sequence similarities (Smith-Waterman scores) against kinases SAR similarity, (**C**) compound-compound (GPCR ligands) structural similarities (Tanimoto scores) against the GPCR ligands SAR similarities and (**D**) compound-compound (kinase ligands) structural similarity (Tanimoto scores) against kinase ligands SAR similarities.
doi:10.1371/journal.pone.0041064.g002

the initial resource located on popular proteins can marginally improve the predictive accuracy of NBI method.

Edge Weighted Network-based Inference (EWNBI). In this study, the edges of CPI network are weighted by the potency (x_{ij}) of binding affinity (K_i) or inhibitory activity (IC_{50}) of the real physical interactions among the chemical and protein nodes (Figure 1D). As given in Table 2, the performance of EWNBI was marginally worse than the unweighted NBI, which is broadly consistent with the strength of the weak ties hypothesis in biochemical network [35].

Role of Weak Chemical-Protein Interactions. To further explore the role of weak interactions in CPI bipartite network, we introduce an exponentλ:

$$a'^{\lambda}_{ij} = \begin{cases} x^{\lambda}_{ij} & K_i(IC_{50}) \le 10\mu M \\ 0 & K_i(IC_{50}) > 10\mu M \end{cases} \quad (5)$$

In EWNBI, when a node i allocates its resource to two nodes p and q, the ratio of resource p and q received is $a^{\lambda}_{ip} \Big/ a^{\lambda}_{iq}$. When $\lambda = 0$, it is the unweighted NBI method; when $\lambda = 1$, it is the EWNBI method. When $\lambda > 1$, it positively strengthens the weighted value of strong CPI edges (high potency between chemical and protein nodes), while $0 < \lambda < 1$ positively strengthens the weighted value of weak CPI edges (low potency between chemical and protein nodes). Otherwise, a negative λ gives the negative effects.

As shown in Figure 4, the AUC increases by λ increasing when$\lambda < 0$. The AUC decreases by λ increasing when$\lambda > 0$. The highest AUC were yielded when $\lambda = 0.50$ and 0.25 for GPCRs and kinases, respectively. The results indicated that the weak interactions could play an important role in CPI prediction in the real weighted CPI network, which is in agreement with the weak ties hypothesis in some real network, such as US air transportation network [36], the neural network of the nematode worm C. elegans [36], the co-authorship network [36], social networks [37] and biochemical network [35] etc. Although it is well-known that weak links hypothesis is very important for complex network, this result is the first confirmation in the real CPI network.

Table 2. The performance of the test set of GPCRs and kinases using different methods by 10 simulation times test of 10-fold cross validation.

Target	Methods	C_i (P_a, P_b, …, P_m)			P_j (C_a, C_b, …, C_n)		
		R	ER	AUC	R	ER	AUC
GPCRs	NBI	0.969±0.004*	18.8±0.086	0.981±0.001	0.285±0.022	270.5±21.1	0.972±0.002
	NWNBI	0.974±0.004	18.9±0.070	0.981±0.001	0.285±0.022	270.5±21.1	0.972±0.002
	EWNBI	0.970±0.004	18.8±0.072	0.981±0.001	0.283±0.028	268.4±26.6	0.972±0.002
	DBSI-T	0.488±0.014	9.48±0.263	0.885±0.002	0.215±0.037	203.7±35.1	0.885±0.004
	DBSI-C	0.458±0.011	8.88±0.213	0.879±0.002	0.159±0.042	151.0±39.6	0.874±0.004
	DBSI-F	0.476±0.013	9.23±0.260	0.880±0.002	0.158±0.040	149.9±37.9	0.874±0.004
	DBSI-R	0.427±0.011	8.27±0.222	0.879±0.002	0.169±0.035	160.1±33.1	0.874±0.004
	TBSI	0.907±0.003	17.6±0.064	0.969±0.001	0.035±0.014	33.51±13.1	0.570±0.007
Kinases	NBI	0.863±0.007	35.5±0.302	0.976±0.002	0.380±0.022	215.0±12.4	0.958±0.001
	NWNBI	0.877±0.007	36.1±0.294	0.977±0.002	0.380±0.022	215.0±12.4	0.958±0.001
	EWNBI	0.866±0.010	35.7±0.397	0.976±0.002	0.360±0.025	203.8±13.9	0.955±0.002
	DBSI-T	0.326±0.015	13.4±0.627	0.878±0.003	0.205±0.022	115.8±12.7	0.846±0.006
	DBSI-C	0.303±0.016	12.5±0.642	0.872±0.003	0.141±0.016	79.8±9.3	0.826±0.007
	DBSI-F	0.273±0.011	11.3±0.465	0.872±0.003	0.137±0.016	77.7±9.0	0.825±0.007
	DBSI-R	0.280±0.014	11.5±0.587	0.872±0.003	0.140±0.016	78.9±9.3	0.827±0.007
	TBSI	0.645±0.007	26.6±0.289	0.908±0.005	0.061±0.011	34.5±6.0	0.660±0.008

All performances were evaluated based on top 5 predicted lists. NBI, network-based inference; NWNBI, node weighted network-based inference; EWNBI, edge weighted network-based inference; DBSI-T, drug-based similarity inference with Tanimoto similarity score; DBSI-C, DBSI with Cosine similarity score; DBSI-F, DBSI with Forbes similarity score; DBSI-R, DBSI with Russell-rao similarity score; TBSI, target-based similarity inference; R, recall; ER, recall enhancement; AUC, the area under the receiver operating characteristic curve; C_i (P_a, P_b, …, P_m) represents the prioritization of new targets for a given chemical; P_j (C_a, C_b, …, C_n) represents the prioritization of new chemicals for a given protein. *The standard deviation of the performance measured by 10 independent simulation times test of 10-fold cross validation.
doi:10.1371/journal.pone.0041064.t002

Prediction of Novel Chemical-Protein Interactions

Although the NBI method can yield high predictive performance, there was a defect that the NBI method cannot predict general chemicals or proteins which did not have any initial links in the training set [26]. In this study, we resolved this bottleneck by integrating the NBI and DBSI methods. Ninety-two novel FDA approved and experimental drugs targeting 46 known GPCRs, and 188 novel approved and experimental drugs targeting 28 known kinases (designated as the external validation sets in Table S3) were collected from DrugBank [38] and KEGG [39], which CPI pairs did not include in the training set (Table 1). Before prioritizing new candidate proteins for a novel chemical (designated compound **A**) using NBI method, we constructed a new initial virtual resources for compound **A** as follows: (i) Calculate the Tanimoto similarity between compound **A** and each compound in the training set; (ii) Displace the topology CPI links of a compound with the highest Tanimoto similarity score in the training set for compound **A**; (iii) Then the candidate proteins were prioritized for compound **A** using the new constructed virtual CPI bipartite network. As summarized in Table 3, the reasonable predictive accuracies were yielded. For GPCRs, the AUC value of 0.77 was yielded when prioritizing candidate targets to a given novel drug using the NBI method, which higher than 0.74 using DBSI method. For 188 drugs of kinase, the AUC value of NBI method was about 0.83, which was marginally lower than DBSI method. The possible reason is that the CPI network of kinases was too sparse, as the average degree of 188 drugs of kinase was only 1.07 (Table 1). In order to assess the reliability of the gold standard data to determine whether the good results might be based on very similar homologous relationships between compounds and similar compounds, we re-evaluated the generalization ability of our methods based on the new validation set after removing 50% high similar compounds with top Tanimoto scores using MACCS keys on the original external validation set (Table S3). As showed in Table S4, the reasonable high performance was also yielded for the new validation set after removing high similar compounds.

Cast Studies

In order to test the real predictive ability of our method, we prioritized all candidate CPIs for known ligands or proteins using the unweighted NBI method by combining the training sets and external validation sets. About 183 thousands of candidate CPI pairs among 4833 known ligands (including 139 FDA approved or experimental drugs) and 97 GPCRs were predicted. About 415 thousands of candidate CPI pairs among 3015 known ligands (including 267 FDA approved or experimental drugs) and 206 kinases were predicted. All predicted CPI lists can be downloaded from web sites: http://www.lmmd.org/database/cpi/ for further experimental investigation. Two known and predicted CPI bipartite networks were constructed using Cytoscape (http://www.cytoscape.org/) in Figures 5 and 6. Due to space limit, we only investigated the predicted targets for five known drugs, namely imatinib, dasatinib, sertindole, olanzapine and ziprasidone. And the molecular hypotheses and experimental evidences of predictions were provided (Table S5).

Imatinib (DB00619) is an ATP-competitive selective inhibitor of Bcr-Abl used to treat chronic myelogenous leukemia (CML), gastrointestinal stromal tumors and a number of other malignancies. In ChEMBL, imatinib (Compound_ID 7083) targeted 9 known kinases, namely $hABL_1$, $hABL_2$, $hPDGFR_a$, $hPDGFR_b$, hLCK, hKIT, hLYN, $hCSF_{1R}$ and hSYK with IC_{50} less than

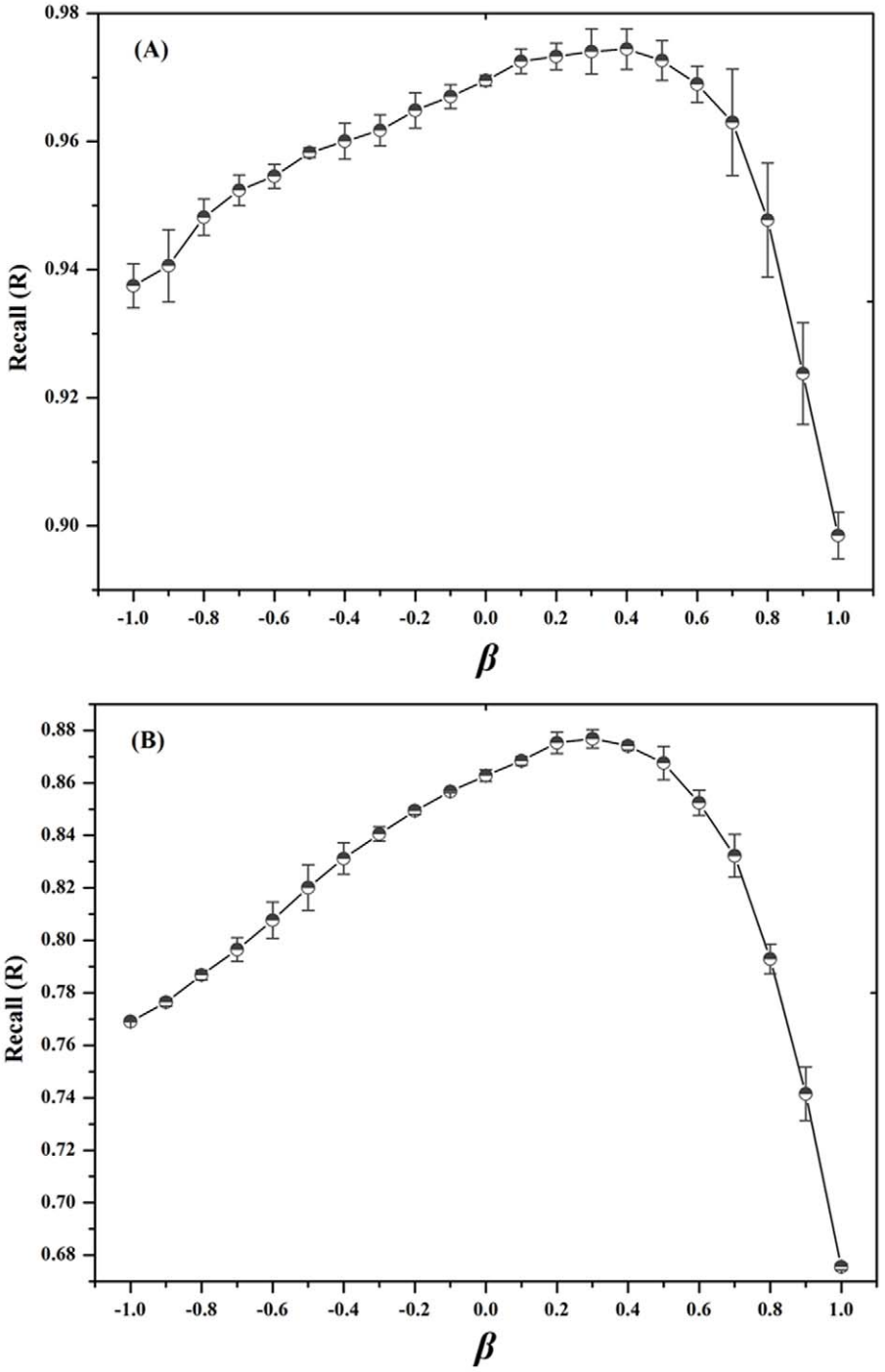

Figure 3. Recall metric of the parameter β on the node weighted network-based inference method for test set when assessed the top five predicted candidate lists. The recall reaches its maximum value at about 0.4 and 0.3 for GPCRs (**A**) and kinases (**B**), respectively. The error bars denote the standard deviation by 10 times independent simulation test.
doi:10.1371/journal.pone.0041064.g003

10 μM (Figure 5) [27]. As given in Table S5, among top 16 predicted kinases of imatinib, 9 ones were predicted correctly with a successful hit rate of 56.3%. Interestingly, seven new targets, namely $hVEGFR_1$, $hVEGFR_2$, hSRC, hEGFR, $hFGFR_1$, $hFLT_3$, and $hTIE_2$, were also predicted for imatinib with high scores. Among them, $hVEGFR_2$ was predicted with the highest score at 0.812. Deininger *et al.* reported that the IC_{50} values of imatinib were 31.2 μM, 19.5 μM and 10.7 μM for $hFGFR_1$, $hVEGFR_1$ and $hVEGFR_2$, respectively [40]. Our prediction was consistent with literatures.

Dasatinib (DB01254) is a novel oral dual, multi-target tyrosine kinase inhibitor, which was approved for chronic myelogenous leukemia treatment. In ChEMBL, dasatinib (Compound_ID 12304) targeted 19 known kinases, namely $hVEGFR_2$, hp38a, hp38b, hp38d, hp38g, hSRC, hEGFR, $hPDGFR_b$, $hFGFR_1$, hLCK, hKIT, $hABL_1$, $hHER_2$, $hMEK_1$, $hMEK_2$, hFYN, $hYES_1$,

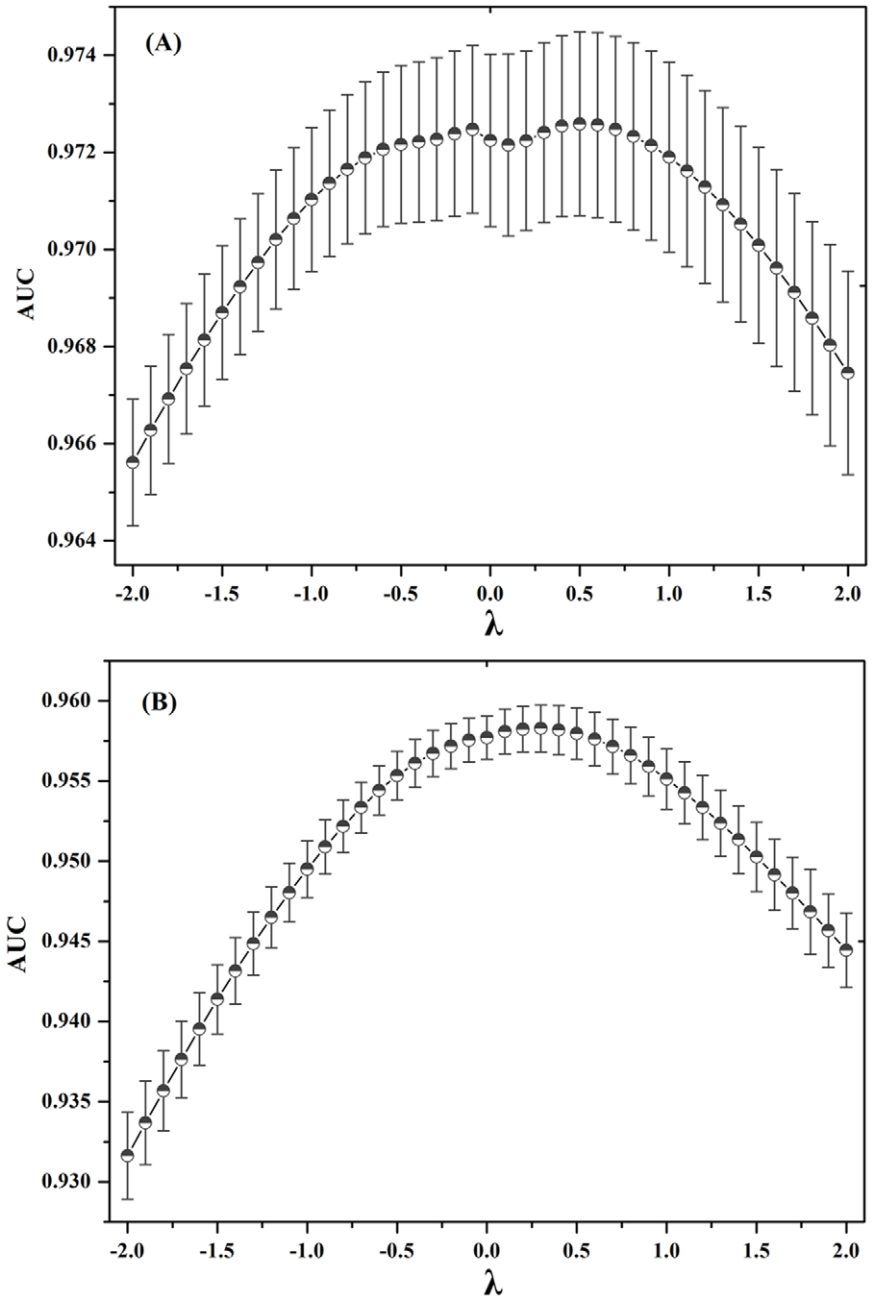

Figure 4. Analysis of the role of weak chemical-protein interactions by exponent λ. When $\lambda = 0$, it is unweighted NBI method; when $\lambda = 1$, it is the EWNBI method. When $\lambda > 1$, it positively$\lambda > 1$ strengthens the weighted value of strong CPI edges, while $0 < \lambda < 10 < \lambda < 1$ positively$\lambda > 1$ strengthens the weighted value of weak CPI edges. Otherwise, a negativeλ will give the negative effects. The area under receiver operating characteristic curve (AUC) was yielded for test set by simulation 10 times test, the error bar denotes the standard deviation. GPCRs (**A**) and kinases (**B**).
doi:10.1371/journal.pone.0041064.g004

$hCSF_{1R}$ and $hEPHA_2$ with the IC_{50} value less than 10 μM. As given in Table S5, on top 27 predicted kinases of dasatinib, 19 targets were predicted correctly with a hit successful rate of 70.4%. Seven new kinases of $hVEGFR_1$, $hVEGFR_3$, $hTIE_2$, $hFLT_3$, $hPDGFR_a$, $hRAF_{1R}$, $hABL_2$, and $hHER_4$ were predicted to bind with dasatinib with high scores. Lombardo *et al.* demonstrated that dasatinib inhibit PDGFR *in vitro* with an IC_{50} of 28 nM [41]. Chen *et al.* reported that dasatinib is a potent inhibitor of PDGFR *via* cell-based assay [42]. Quintas-Cardama *et al.* reported that dasatinib effectively inhibited several SRC family kinases, including SRC ($IC_{50} = 0.55$ nM), LCK (($IC_{50} = 1.1$ nM), FYN ($IC_{50} = 0.2$ nM) and YES ($IC_{50} = 0.41$ nM) [43]. The data indicated that our predicted results are in agreement with literatures.

Sertindole (DB06144) is an oral antipsychotic drug targeted with dopamine D_2, serotonin $5\text{-}HT_{2A}$ and $5\text{-}HT_{2C}$, and α_1-adrenoreceptors. The clinical trails have confirmed that sertindole is effective at a low dopamine D_2 occupancy level. In the ChEMBL, sertindole (Compound_ID 85092) targeted 15 known GPCRs, namely DRD_1, DRD_2, DRD_3, DRD_4, A_{1AA}, A_{1AB}, A_{1AD},

Table 3. The performance of difference inference methods in the external validation set of GPCRs and kinases.

Tragets	Methods	C_i (P_a, P_b, …, P_m)			P_j (C_a, C_b, …, C_n)		
		R	ER	AUC	R	ER	AUC
GPCRs	NBI	0.535	2.60	0.769	0.684	3.15	0.693
	NWNBI	0.559	2.72	0.756	0.684	3.15	0.693
	EWNBI	0.561	2.72	0.764	0.697	3.21	0.691
	DBSI-T	0.470	2.28	0.743	0.603	2.77	0.685
	DBSI-C	0.472	2.29	0.739	0.604	2.78	0.684
	DBSI-F	0.473	2.29	0.739	0.612	2.82	0.686
	DBSI-R	0.473	2.30	0.741	0.610	2.81	0.683
	TBSI	0.342	1.66	0.639	0.361	1.66	0.593
Kinases	NBI	0.502	5.17	0.828	0.222	2.09	0.607
	NWNBI	0.427	4.40	0.812	0.222	2.09	0.607
	EWNBI	0.459	4.73	0.821	0.159	1.50	0.597
	DBSI-T	0.594	6.11	0.847	0.188	1.77	0.573
	DBSI-C	0.588	6.06	0.847	0.148	1.39	0.564
	DBSI-F	0.595	6.12	0.846	0.142	1.33	0.563
	DBSI-R	0.583	6.00	0.846	0.146	1.38	0.563
	TBSI	0.061	0.62	0.326	0.082	0.775	0.510

All performances were evaluated based on top 20 predicted lists. NBI, network-based inference; NWNBI, node weighted network-based inference; EWNBI, edge weighted network-based inference; DBSI-T, drug-based similarity inference with Tanimoto similarity score; DBSI-C, DBSI with Cosine similarity score; DBSI-F, DBSI with Forbes similarity score; DBSI-R, DBSI with Russell-rao similarity score; TBSI, target-based similarity inference; R, recall; ER, recall enhancement; AUC, the area under the receiver operating characteristic curve; C_i (P_a, P_b, …, P_m) represents the prioritization of new targets for a given chemical; P_j (C_a, C_b, …, C_n) represents the prioritization of new chemicals for a given protein.
doi:10.1371/journal.pone.0041064.t003

A_{2AA}, A_{2AC}, $5HT_{1A}$, $5HT_{1B}$, $5HT_{2A}$, $5HT_{2C}$, A_{2AB} and HRH_1 with K_i value less than 10 μM (Figure 6). As given in Table S5, in top 16 predicted GPCRs of sertindole, 15 receptors were predicted correctly with a hit successful rate of 93.8%. Sertindole was first marketed in 1996 in several European countries. However, it was withdraw two years later because of numerous cardiac adverse effects such as QTc prolongation and the UK database adverse drug reactions information tracking reported that the rate of arrhythmias or sudden death was almost 10-times greater for sertindole than for olanzapine and risperidone [44]. The molecular mechanism of side effects of sertindole was unknown. As given in Figure 6, sertindole was predicted to bind with hB_{1AR} and hB_{2AR}, which were consistent with literature [45].

Olanzapine (DB00334) approved in 1996, is an atypical antipsychotic agent, which is used to treat both negative and positive symptoms of schizophrenia, acute mania with bipolar disorder, agitation, and psychotic symptoms in dementia [38,46]. Olanzapine mainly targeted with dopamine, histamine H_1, muscarinic, 5-HT_2 and α_1-adrenoreceptors with high binding affinities. In the ChEMBL, olanzapine promiscuously targeted 22 known GPCRs, namely $hDRD_1$, $hDRD_2$, $hDRD_3$, $hDRD_4$, $hDRD_5$, $hCHRM_1$, $hCHRM_2$, $hCHRM_3$, $hCHRM_4$, $hCHRM_5$, hA_{1AB}, hA_{1AA}, hA_{1AD}, hA_{2AA}, hA_{2AB}, hB_{1AR}, hB_{2AR}, hB_{3AR}, $h5HT_{2A}$, $h5HT_{2C}$, $h5HT_6$ and $hHRH_1$ (Figure 6). As given in Table S5, in top 26 predicted GPCRs of olanzapine, 22 targets were predicted correctly with a hit successful rate of 84.6%. The receptors of hA_{2AC}, $h5HT_{1A}$, $h5HT_{1B}$, $h5HT_{1D}$, $h5HT_{1E}$, $h5HT_{2B}$, $h5HT_7$, $hHRH_2$ and $hHRH_3$ were predicted to have novel interactions with olanzapine (Figure 6). Recently, several results reported that olanzapine can bind with the receptors of $h5HT_{1A}$, $h5HT_{1B}$, $h5HT_{1D}$, $h5HT_{1E}$ and $h5HT_7$ with high binding affinities [45,47].

Ziprasidone (DB00246) is a selective monoaminergic antagonist with high affinity for the serotonin Type ($5HT_{1A}$, $5HT_2$), dopamine D_2 and H_1 histaminergic receptors. It is a psychotropic agent indicated for the treatment of schizophrenia. In the ChEMBL, ziprasidone (Compound_ID 89351) targeted 16 known GPCRs, namely $hDRD_1$, $hDRD_2$, $hDRD_3$, $hDRD_4$, hA_{1AA}, hA_{1AD}, hA_{2AA}, hA_{2AB}, hA_{2AC}, $h5HT_{1A}$, $h5HT_{2A}$, $h5HT_{2B}$, $h5HT_{2C}$, $h5HT_6$, $hHRH_1$ and $hCHRM_1$ with K_i value less than 10 μM (Figure 6). As given in Table S5, in top 20 predicted GPCRs of ziprasidone, 16 targets were predicted correctly with a hit successful rate of 80%. The receptors of $hDRD_5$, $h5HT_{1B}$, $h5HT_{1D}$, $h5HT_7$, $hCHRM_2$, $hCHRM_3$, $hCHRM_4$, $hCHRM_6$, hB_{1AR} and hB_{2AR} were predicted to have new indications with ziprasidone. Recently, several results reported that ziprasidone can bind with the receptors of $h5HT_{1B}$, $h5HT_{1D}$, $h5HT_7$, $hCHRM_2$, $hCHRM_3$, $hCHRM_4$, $hCHRM_6$, hB_{1AR} and hB_{2AR} with the high binding affinities [45,47], which demonstrated the feasibility of our methods to prioritize new target to known drugs.

Discussion

Potential Application of Our Methods

Herein, we systematically investigated the utility of unweighted and weighted network-based inference method in prediction of new targets for old drugs or general ligands. The proposed method achieved the AUC was about 0.98 and 0.83 for the test set and the external validation set, respectively. Today, the increased availability of large scale open access resources on bioactivities of small molecules has a significant impact on pharmacology facilitated [48]. Therefore, our methods could provide a fast and effective strategy to digest the vast amounts of data for CPI prediction and drug repositioning.

The method proposed in this study fall within the scope of the emerging field of systems pharmacology [1]. Recently, systems pharmacology approaches have been applied successfully to various problems, such as drug repositioning [1,26]. Herein, we extended our previous work on developing two different weighted NBI, namely EWNBI and NWNBI for CPI prediction and drug repositioning. We found that NWNBI method was marginally higher than NBI with an appropriate parameter optimization. And the weak interactions hypothesis was first proposed in CPI network by EWNBI method. To our knowledge, our method could be used in several biological relevant directions, such as gene-disease association prediction [49], drug-diseases association prediction (drug adverse events prediction) [50] etc. by integrating meta-biochemical networks in the further.

Polypharmacology of Ligands

The resistance of anti-cancer drugs is a large challenge for cancer therapeutics [51]. Overcoming the resistance mechanisms may require targeting tumor cells at promiscuous levels, through either single drugs binding with the multi-targets or cocktails of several highly selective inhibitors [52]. A big bottleneck for the cancer research community is how to decipher chemical-protein interactome, how to optimize the best combinations of targets and then prioritize those combinations for clinical testing. The polypharmacology of kinase ligands was encouraged by our results (Figure 5) that the emerging class of well-tolerated kinase inhibitors of imatinib and dasatinib, exhibit the multi-target on kinases and are less selective than initially findings [53]. Today, more than 800

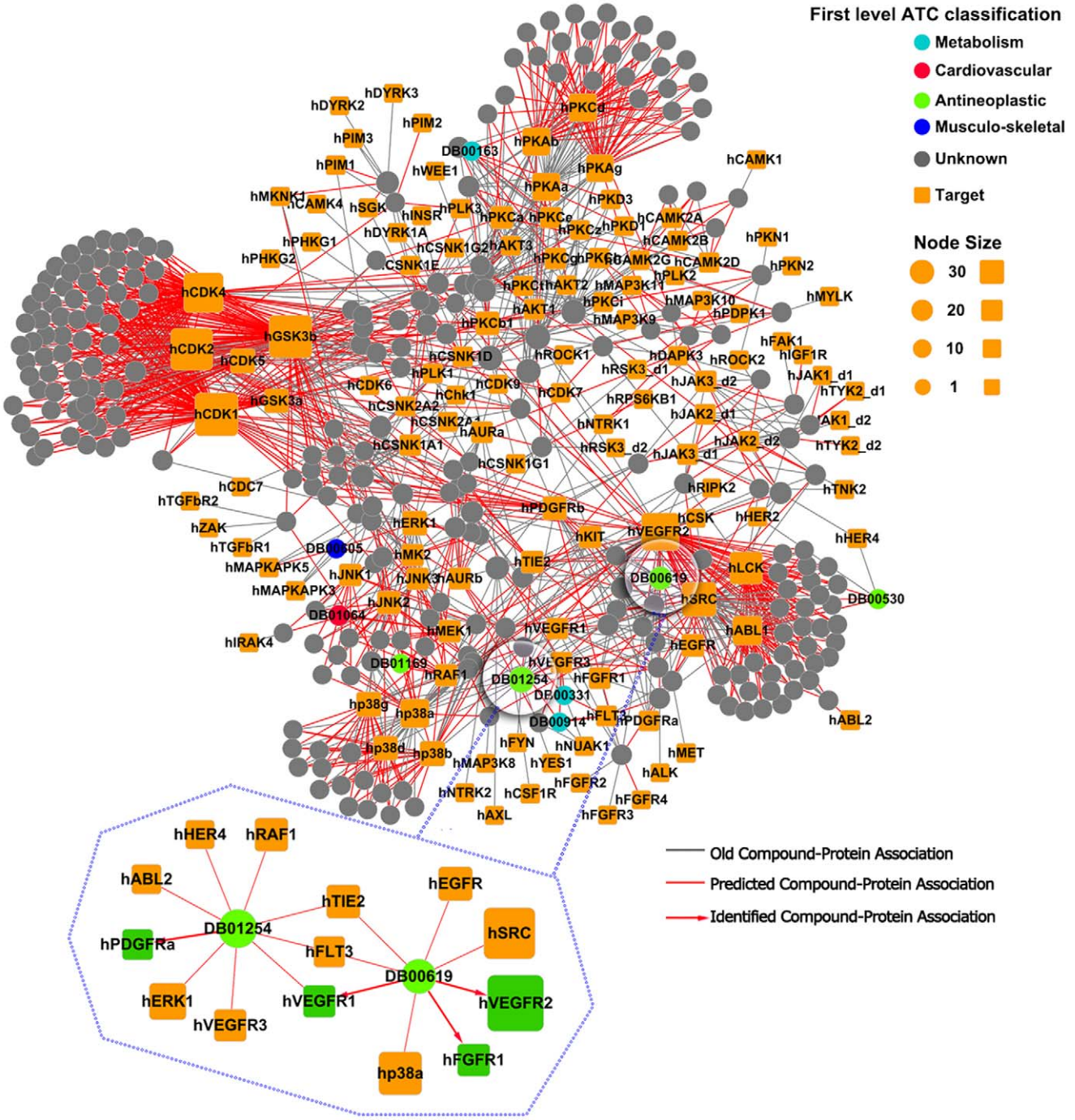

Figure 5. Discovered chemical-protein interactions (CPI) bipartite networks among 267 FDA approved or experimental drugs and 130 kinases. Circle and square nodes correspond to drugs and kinases, respectively. A gray line represents the old CPI annotated in the DrugBank and KEGG. The red line represents the predicted CPI. The red arrow line represents the new predicted CPI which is validated by literatures. The size of the drug node is the fraction of the number of targets that the drug linked. The size of the target node is the fraction of the number of drugs that the target linked. Color codes are given in the legend. Drug nodes (circles) are colored according to their Anatomical Therapeutic Chemical Classification. This graph and Figure 6 were prepared by Cytoscape (http://www.cytoscape.org/).
doi:10.1371/journal.pone.0041064.g005

drugs are in clinical development for cancer indications and the current success rate in bringing drugs to the markets remains only in the range of 5–8% [2]. Most of drugs which selectively targeted kinases with high potency *in vitro* models are failure in the late stage of clinical trails due to side effects and lacking *in vivo* activities [54,55]. As shown in Figure 5, most of experimental drugs of kinases (gray circle nodes) in DrugBank were predicted to have a significant promiscuity. These results could be useful for finding new usages of some failure of drugs and multi-target anticancer drugs design. For example, imatinib was initially approved for the treatment of CML, but it was tested in five patients with hypereosinophilic syndrome [56].

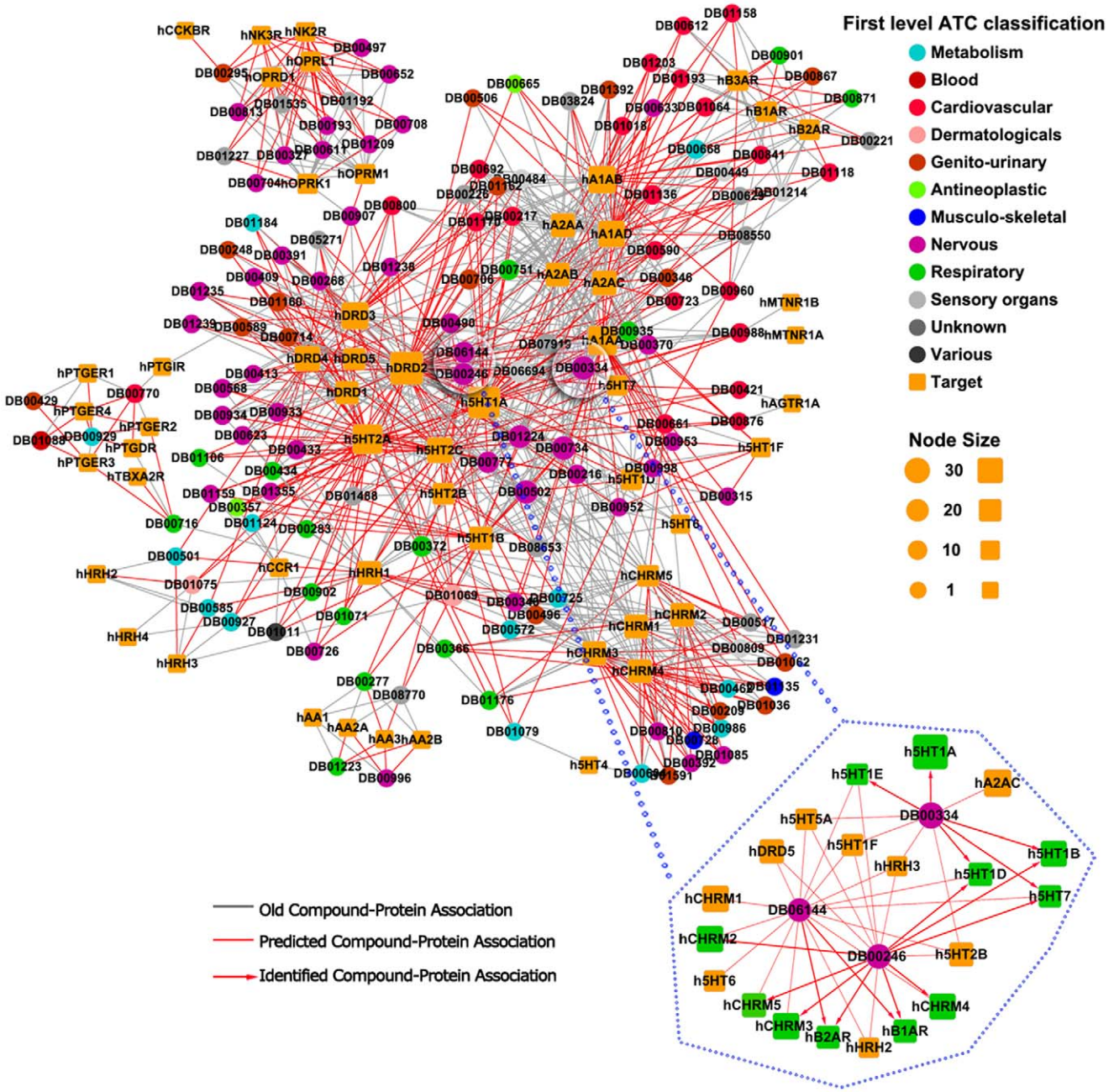

Figure 6. Discovered chemical-protein interaction (CPI) bipartite network among 139 FDA approved or experimental drugs and 55 GPCRs (Table S4). Circle and square nodes correspond to drugs and GPCRs, respectively. The definition of nodes and edges were given in the caption of Figure 5.
doi:10.1371/journal.pone.0041064.g006

The features of polypharmacology are not restricted to the kinases inhibitors. Anti-psychotics drugs, such as sertindole, olanzapine and ziprasidone also promiscuously targeted with GPCRs rather than individual one. As given in Figure 6, the therapeutic effects of sertindole, olanzapine and ziprasidone mainly targeted three dopaminergic (D_1, D_2 and D_3) and three serotonin-ergic (5-HT_{1A}, 5-HT_{1D} and 5-HT_{2A}) receptors. But they often lead to several side effects by binding with adrenergic and histaminergic receptors, such as QTc prolongation [45]. In this study, it is worth acknowledged that NBI methods can effectively help to prioritize the new candidate CPI, decipher potential molecular mechanism of off-targets and drug repositioning.

Weak Interactions between Chemicals and Proteins

Herein, for the first time we identified the evidence of weak interactions in CPI network. In fact, multiple weak interactions cannot be ignored in polypharmacological actions [57]. It is estimated that most (more than 80%) of the cellular proteins, signaling and transcriptional networks are in a low-affinity or transient "weak linkage" with each other [58]. Weak physical interactions with low binding affinity play critical roles in molecular recognition among biological systems, from the classic example of protein folding to recent discoveries in metabolism, gene regulation and signal transduction [59]. For example, the binding affinity between enzyme and the alternative substrate is usually low [60].

The hypothesis of weak interactions for drug therapeutics had been applied for more than two thousand years in Chinese Traditional Medicine [61]. A drug with low affinity and multi-target may have high therapeutic value with fewer side effects than one with high affinity and single target. For example, sorafenib was designed as a potent nanomolar (nM) inhibitor of BRAF which a protein implicated in the survival of melanoma cells. Unfortunately, it failed in clinical trial due to its low anti-melanoma efficacy [62]. In contrast to, low affinity and multi-target noncompetitive NMDA receptor antagonists developed for treatment of Alzheimer's disease, may have fewer side effects than some high affinity and single target drugs [63]. The detailed assessment of weak CPI interaction is a hot topic in drug discovery and complex network, but it was beyond the range of this article. Our groups are actively investigating this important issue.

Conclusions

In summary, we proposed two different weighted NBI methods for CPI prediction. The high performance was yielded using our methods. Comparing with conventional ligand and receptor-based methods, NBI method only used CPI network topology similarity by simultaneously exploiting both topological and functional modularity to prioritize new targets for a given drug or prioritize new drugs for a given target, which did not need any 2D or 3D structural information of targets and drugs. Our methods will generate a set of predicted candidate miss linked CPI. The biologist can then follow up on the new high scoring CPI for further experimental assay. Therefore, our methods open new avenue for CPI identification.

The weak links hypothesis had been proposed in several biochemical networks and social network etc. In this study, the weak links hypothesis in CPI network was first proposed by EWNBI method. Enhance and diminish stronger or weaker CPI edges all decreased the predict accuracy. The maximum predictive accuracy was yielded when stronger and weaker CPI edges achieved a balance. These computational polypharmacology perspectives could let people beef up efforts for CPI prediction and drug repositioning.

Supporting Information

Figure S1 The degree distributions of chemical and protein (GPCRs and Kinases) nodes in two comprehensive chemical-protein interactions bipartite networks.
(TIF)

Figure S2 The heat maps of the chemical similarities and protein sequence similarities. A) Tanimoto similarity of 4,741 GPCRs ligands, **B**) Tanimoto similarity of 2,827 kinase ligands, **C**) Genomic sequence Smith-Waterman similarity of 97 GPCRs, **D**) Genomic sequence Smith-Waterman similarity of 206 kinases.
(TIF)

Table S1 The detailed description of data sets using in this study. The Compound ID, Target ID, SIMLES, K_i and IC_{50} value of 17,100 chemical-protein interaction pairs among 4,741 compounds and 97 G protein-coupled receptors (GPCRs), and 13,600 CPI pairs among 2,827 compounds and 206 kinases.
(XLS)

Table S2 The sequences in FASTA format of 97 G protein-coupled receptors and 206 kinases extracted from ChEMBL database.
(DOC)

Table S3 The detailed description of the external validation set. The external validation set of 92 novel FDA approved and experimental drugs targeting 46 GPCRs, and 188 novel approved and experimental drugs targeting 28 kinases collected from DrugBank and KEGG.
(XLS)

Table S4 The performance of difference inference methods on the new external validation set. The new external validation set was constructed after removing 50% high similar compounds with top Tanimoto similarity using MACCS keys on the original external validation set (Table S3) of GPCRs and kinases.
(PDF)

Table S5 The molecular hypothesis, experimental evidence and predicted target list for five known drugs, namely imatinib, dasatinib, sertindole, olanzapine and ziprasidone.
(XLS)

Author Contributions

Conceived and designed the experiments: YT FC. Performed the experiments: FC YZ. Analyzed the data: FC YZ WL GL. Contributed reagents/materials/analysis tools: YT FC. Wrote the paper: FC YT.

References

1. Hopkins AL (2008) Network pharmacology: the next paradigm in drug discovery. Nat Chem Biol 4: 682–690.
2. Schilsky RL, Allen J, Benner J, Sigal E, McClellan M (2010) Commentary: tackling the challenges of developing targeted therapies for cancer. Oncologist 15: 484–487.
3. Yildirim MA, Goh KI, Cusick ME, Barabasi AL, Vidal M (2007) Drug-target network. Nat Biotechnol 25: 1119–1126.
4. Uliana SR, Barcinski MA (2009) Repurposing for neglected diseases. Science 326: 935.
5. Ashburn TT, Thor KB (2004) Drug repositioning: identifying and developing new uses for existing drugs. Nat Rev Drug Discov 3: 673–683.
6. Gonzalez-Diaz H, Prado-Prado F, Garcia-Mera X, Alonso N, Abeijon P, et al. (2011) MIND-BEST: Web Server for Drugs and Target Discovery; Design, Synthesis, and Assay of MAO-B Inhibitors and Theoretical-Experimental Study of G3PDH Protein from Trichomonas gallinae. J Proteome Res 10: 1698–1718.
7. Li H, Gao Z, Kang L, Zhang H, Yang K, et al. (2006) TarFisDock: a web server for identifying drug targets with docking approach. Nucleic Acids Res 34: W219–224.
8. Luo H, Chen J, Shi L, Mikailov M, Zhu H, et al. (2011) DRAR-CPI: a server for identifying drug repositioning potential and adverse drug reactions via the chemical-protein interactome. Nucleic Acids Res 39: W492–498.
9. Yamanishi Y, Araki M, Gutteridge A, Honda W, Kanehisa M (2008) Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. Bioinformatics 24: i232–240.
10. Yabuuchi H, Niijima S, Takematsu H, Ida T, Hirokawa T, et al. (2011) Analysis of multiple compound-protein interactions reveals novel bioactive molecules. Mol Syst Biol 7: 472.
11. Wang F, Liu D, Wang H, Luo C, Zheng M, et al. (2011) Computational screening for active compounds targeting protein sequences: methodology and experimental validation. J Chem Inf Model 51: 2821–2828.
12. van Westen GJ, Wegner J, Ijzerman AP, van Vlijmen HW, Bender A (2011) Proteochemometric modeling as a tool to design selective compounds and for extrapolating to novel targets. Med Chem Commun 2: 16–30.
13. Iorio F, Bosotti R, Scacheri E, Belcastro V, Mithbaokar P, et al. (2010) Discovery of drug mode of action and drug repositioning from transcriptional responses. Proc Natl Acad Sci U S A 107: 14621–14626.
14. Dudley JT, Sirota M, Shenoy M, Pai RK, Roedder S, et al. (2011) Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease. Sci Transl Med 3: 96ra76.
15. Barabasi AL, Gulbahce N, Loscalzo J (2011) Network medicine: a network-based approach to human disease. Nat Rev Genet 12: 56–68.
16. Zhang SD, Lu WQ, Liu XF, Diao YY, Bai F, et al. (2011) Fast and effective identification of the bioactive compounds and their targets form the medicinal

plants via computational chemical biology approach. Med Chem Comm 2: 471–477.
17. Campillos M, Kuhn M, Gavin AC, Jensen LJ, Bork P (2008) Drug target identification using side-effect similarity. Science 321: 263–266.
18. Faulon JL, Misra M, Martin S, Sale K, Sapra R (2008) Genome scale enzyme-metabolite and drug-target interaction predictions using the signature molecular descriptor. Bioinformatics 24: 225–233.
19. Jacob L, Vert JP (2008) Protein-ligand interaction prediction: an improved chemogenomics approach. Bioinformatics 24: 2149–2156.
20. Bleakley K, Yamanishi Y (2009) Supervised prediction of drug-target interactions using bipartite local models. Bioinformatics 25: 2397–2403.
21. Yamanishi Y, Kotera M, Kanehisa M, Goto S (2010) Drug-target interaction prediction from chemical, genomic and pharmacological data in an integrated framework. Bioinformatics 26: i246–254.
22. van Laarhoven T, Nabuurs SB, Marchiori E (2011) Gaussian interaction profile kernels for predicting drug-target interaction. Bioinformatics 27: 3036–3043.
23. Cheng FX, Zhou YD, Li J, Li WH, Liu GX, et al. (2012) Prediction of Chemical-Protein Interactions: Multitarget-QSAR versus Computational Chemogenomic Methods. Mol BioSyst, in press, doi: 101039/C2MB25110H.
24. Zhou T, Kuscsik Z, Liu JG, Medo M, Wakeling JR, et al. (2010) Solving the apparent diversity-accuracy dilemma of recommender systems. Proc Natl Acad Sci USA 107: 4511–4515.
25. Zhou T, Su RQ, Liu RR, Jiang LL, Wang BH, et al. (2009) Accurate and diverse recommendations via eliminating redundant correlations. New J Phys 11: 123008.
26. Cheng FX, Liu C, Jiang J, Lu WQ, Li WH, et al. (2012) Prediction of Drug-Target Interactions and Drug Repositioning via Network-Based Inference. PLoS Comput Biol 8: e1002503.
27. Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, et al. (2012) ChEMBL: a large-scale bioactivity database for drug discovery. Nucleic Acids Res, 40: D1100–1107.
28. Open Babel (version 2.3.0). <http://openbabel.org/> (Access Date: Apr. 18, 2010).
29. Willett P (2006) Similarity-based virtual screening using 2D fingerprints. Drug Discov Today 11: 1046–1053.
30. Smith TF, Waterman MS (1981) Identification of common molecular subsequences. J Mol Biol 147: 195–197.
31. Jia C, RR L, D S, Wang B (2008) A new weighting method in network-based recommendation. Physica A 387: 5887–5891.
32. Bamborough P, Drewry D, Harper G, Smith GK, Schneider K (2008) Assessment of chemical coverage of kinome space and its implications for kinase drug discovery. J Med Chem 51: 7898–7914.
33. Varnek A, Tropsha A (2008) Chemoinformatics: An Approach to Virtual Screening. Cambridge, UK: R. Soc. Chem.
34. Sadowski MI, Jones DT (2009) The sequence-structure relationship and protein function prediction. Curr Opin Struct Biol 19: 357–362.
35. Csermely P (2004) Strong links are important, but weak links stabilize them. Trends Biochem Sci 29: 331–334.
36. Linyuan L, Tao Z (2010) Link prediction in weighted networks: The role of weak ties. EPL 89: P18001–P18006.
37. Granovetter M (1973) The strength of weak ties. Am J Sociol 78: 1360–1380.
38. Knox C, Law V, Jewison T, Liu P, Ly S, et al. (2011) DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. Nucleic Acids Res 39: D1035–1041.
39. Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, et al. (2006) From genomics to chemical genomics: new developments in KEGG. Nucleic Acids Res 34: D354–357.
40. Deininger MW, Druker BJ (2003) Specific targeted therapy of chronic myelogenous leukemia with imatinib. Pharmacol Rev 55: 401–423.
41. Lombardo LJ, Lee FY, Chen P, Norris D, Barrish JC, et al. (2004) Discovery of N-(2-chloro-6-methyl-phenyl)-2-(6-(4-(2-hydroxyethyl)-piperazin-1-yl)-2-methylpyrimidin-4- ylamino)thiazole-5-carboxamide (BMS-354825), a dual Src/Abl kinase inhibitor with potent antitumor activity in preclinical assays. J Med Chem 47: 6658–6661.
42. Chen Z, Lee FY, Bhalla KN, Wu J (2006) Potent inhibition of platelet-derived growth factor-induced responses in vascular smooth muscle cells by BMS-354825 (dasatinib). Mol Pharmacol 69: 1527–1533.
43. Quintas-Cardama A, Kantarjian H, Cortes J (2006) Targeting ABL and SRC kinases in chronic myeloid leukemia: experience with dasatinib. Future Oncol 2: 655–665.
44. Lindstrom E, Levander S (2006) Sertindole: efficacy and safety in schizophrenia. Expert Opin Pharmacother 7: 1825–1834.
45. Nasrallah HA (2008) Atypical antipsychotic-induced metabolic side effects: insights from receptor-binding profiles. Mol Psychiatry 13: 27–35.
46. Chen X, Ji ZL, Chen YZ (2002) TTD: Therapeutic Target Database. Nucleic Acids Res 30: 412–415.
47. Zhang JY, Kowal DM, Nawoschik SP, Lou Z, Dunlop J (2006) Distinct functional profiles of aripiprazole and olanzapine at RNA edited human 5-HT2C receptor isoforms. Biochem Pharmacol 71: 521–529.
48. Iskar M, Zeller G, Zhao XM, van Noort V, Bork P (2011) Drug discovery in the age of systems biology: the rise of computational approaches for data integration. Curr Opin Biotechnol. 23: 1–8, doi: 10.1016/j. copbio.2011.11.010.
49. Gottlieb A, Stein GY, Ruppin E, Sharan R (2011) PREDICT: a method for inferring novel drug indications with application to personalized medicine. Mol Syst Biol 7: 496.
50. Cami A, Arnold A, Manzi S, Reis B (2011) Predicting adverse drug events using pharmacological network models. Sci Transl Med 3: 114ra127.
51. Knight ZA, Lin H, Shokat KM (2010) Targeting the cancer kinome through polypharmacology. Nat Rev Cancer 10: 130–137.
52. Sawyers CL (2007) Cancer: mixing cocktails. Nature 449: 993–996.
53. Fabian MA, Biggs WH 3rd, Treiber DK, Atteridge CE, Azimioara MD, et al. (2005) A small molecule-kinase interaction map for clinical kinase inhibitors. Nat Biotechnol 23: 329–336.
54. Overington JP, Al-Lazikani B, Hopkins AL (2006) How many drug targets are there? Nat Rev Drug Discov 5: 993–996.
55. Imming P, Sinning C, Meyer A (2006) Drugs, their targets and the nature and number of drug targets. Nat Rev Drug Discov 5: 821–834.
56. Gleich GJ, Leiferman KM, Pardanani A, Tefferi A, Butterfield JH (2002) Treatment of hypereosinophilic syndrome with imatinib mesilate. Lancet 359: 1577–1578.
57. Xie L, Kinnings SL, Bourne PE (2012) Novel computational approaches to polypharmacology as a means to define responses to individual drugs. Annu Rev Pharmacol Toxicol 52: 361–379.
58. Korcsmaros T, Szalay MS, Bode C, Kovacs IA, Csermely P (2007) How to design multi-target drugs: Target search options in cellular networks. Expert Opin Drug Discovery 2: 799–808.
59. Akitaya T, Seno A, Nakai T, Hazemoto N, Murata S, et al. (2007) Weak interaction induces an ON/OFF switch, whereas strong interaction causes gradual change: folding transition of a long duplex DNA chain by poly-L-lysine. Biomacromolecules 8: 273–278.
60. D'Ari R, Casadesus J (1998) Underground metabolism. Bioessays 20: 181–186.
61. Shen ZX, Shi ZZ, Fang J, Gu BW, Li JM, et al. (2004) All-trans retinoic acid/As2O3 combination yields a high quality remission and survival in newly diagnosed acute promyelocytic leukemia. Proc Natl Acad Sci U S A 101: 5328–5335.
62. Gleeson MP, Hersey A, Montanari D, Overington J (2011) Probing the links between in vitro potency, ADMET and physicochemical parameters. Nat Rev Drug Discov 10: 197–208.
63. Youdim MB, Buccafusco JJ (2005) Multi-functional drugs for various CNS targets in the treatment of neurodegenerative disorders. Trends Pharmacol Sci 26: 27–35.