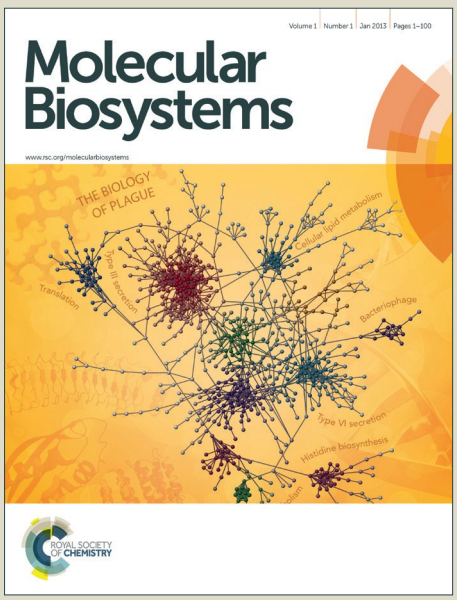
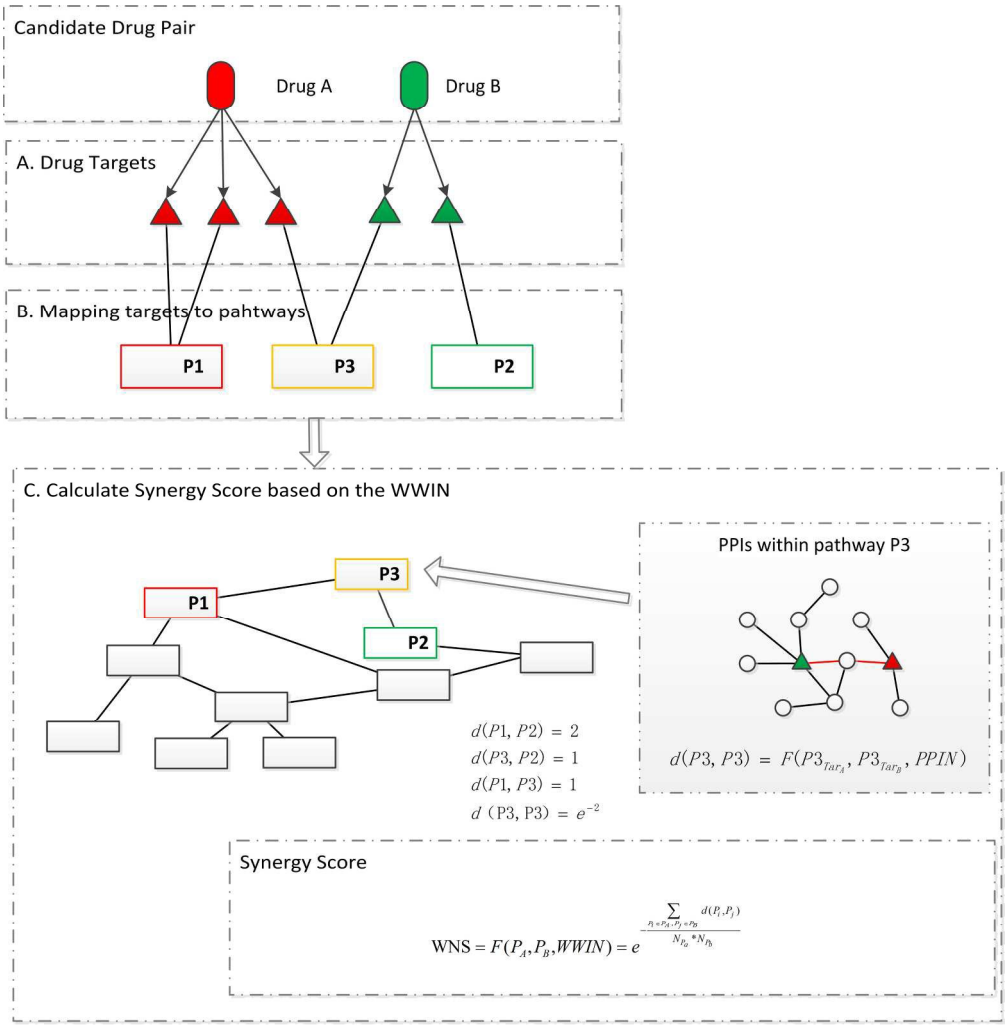
Molecular BioSystems

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/molecularbiosystems



211x215mm (300 x 300 DPI)

Synergy evaluation by a pathway-pathway interaction network: a new way to predict drug combination

Di Chen‡[a], Huamin Zhang‡[b], Peng Lu*[a], Xianli Liu[c] and Hongxin Cao*[d]

[a] Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

[b] Institute of Information on TCM, China Academy of Chinese Medical Sciences, Beijing 100700,China

[c] Institute of Basic Theory of TCM, China Academy of Chinese Medical Sciences, Beijing 100700,China

[d] State Administration of Traditional Chinese Medicine of the People's Republic of China, Beijing 100027, China.

* Corresponding to Peng Lu: peng.lu@ia.ac.cn and Hongxin Cao: caohx@mail.cintcm.ac.cn

‡ Contributed equally

Abstract

Drug combinations have been widely applied to treat complex diseases, like cancer, HIV and cardiovascular diseases. One of the most important characteristics for drug combinations is the synergistic effects among different drugs, that is to say, the combination effects is larger than the sum of individual effects. Although quantitative methods can be utilized to evaluate the synergistic effects based on experimental dose-response data, it is both time and resource consuming to screen all possible combinations by experiment trails. This problem makes it a formidable challenge to recognize synergistic combinations. Various attempts have been put on the prediction of drug synergy by network biology, however, most of them are limited to estimating target associations on the PPI network. Here, we proposed a novel "pathway-pathway interaction" network-based synergy evaluation method to

predict the potential synergistic drug combinations. Comparison with previous target-based methods shows inclusion of systematic pathway-pathway interactions makes this novel method outperform others in predicting drug synergy. Moreover, it can also help to interpret how different drugs in a combination cooperate with each other to implement synergistic therapeutic effect. In general, drugs acting on the same pathway through different targets or drugs regulating a relatively small number of highly-connected pathways are more likely to produce synergistic effects.

Introduction

Drug combination, in which two or more agents are administrated together, is increasingly applied to treat complex diseases like HIV and cancer [1, 2]. One of the most important benefits of drug combination is synergistic effect, which means the combination effects is larger than the sum of individual drugs[3]. The synergistic effect can help to improve therapeutic efficency and reduce side effects or drug resistance by modulating the activity of multiple proteins,thus overcoming the limitation of "one drug,one target" approach against the systematic and multifunctional physiological processes of complex disease[4]. Consequently, drug combination is gradually becoming one promising way to conquer complex diseases.

Experimentally, dose-response based methods like Loewe additivity[5], isobologram[6] and Chou-Talalay method [7] have been devised to evaluate the interactions between drugs in a combination. However, since such methods are applied to quantify drug synergy in a case-by-case way and certain number of experiments are required to obtain the dose-response data for one combination, it will be both time and resource-consuming to screen all possible combinations among the increasingly available drugs. In addition, these dose-response methods cannot provide clues on the latent mechanism.

Owing to the above limitations, some computational approaches, which utilized various drug-related informaion like chemical structure, target, ATC code and side effect, have been applied to predict and explore drug combinations. For example, Li et al. designed a PPI network-based method

termed "NIMS" [8] to evaluate drug synergy by exploring the topological associations between targets of different drugs. Wang et al. discovered that drug combinations tended to target proteins which were closer in the PPI network comparing against random combinations [9]. Zou et al. utilized neighbor communities, which describe the interactions between drug targets and their neighbors in the PPI network, to distinguish synergistic drug combinations [10]. These methods focus on the relations of targets from different drugs. Taking more drug properties into account, Zhao et al. proposed a prediction method by integrating both molecular and pharmacological features[11], where combinations with features enriched by approved ones were identified as novel combinations. Besides, Chen et al. predict drug combination based on chemical interaction, protein interaction and target enrichment of KEGG pathways [12]. Among these computational methods, the network-based ones are widely used. They can not only help predict the potential synergistic drug combinations, but also help to illustrate the latent molecular mechanism of drug synergy from the perspective of mutual interactions. It seems that topology relationships of drug targets on the PPI network are of great importance for estimating drug synergy. Although the PPI network-based methods can be applied to explore promising drug combinations, they can only help understand synergism in terms of target associations. To compensate for the limitation of targets in reflecting biological function, pathway analysis is often conducted to unveil underlying mechanism [9, 13, 14]. In the study of Wang et al., they observed that drug combinations are more likely to modulate functionally related pathways [9]. Likewise, Zou et al.

discovered that drug combinations mainly act on multiple targets of a pathway and its crosstalk pathways [10]. Investigating on some promising signaling pathway modulatory drug combinations, Dent et al. came to the inclusion that rational drug combinations should simultaneously inhibit multiple inter-related pathways [14]. Another extensive investigation on the molecular mechanism of drug combinations pointed out that "pathway analysis can be an effective approach for a more comprehensive assessment of drug combination effects" [15]. Due to the effectiveness in illustrating the mechanism of drug combination, some pathway-based approaches aiming at identifying synergistic drug combinations have been proposed. Most of these approaches are based on the dynamic simulation[16] of specific pathway like the EGFR signaling pathway [17, 18]. However, the incompleteness of kinetics parameters and lots of constraints specified for model utilization often makes it difficult to implement pathway simulations massively, thus limiting the simulations to only several specific pathways. Consequently, to do a more systematical exploration of effective drug combinations with respect to pathways, we first need to build a more comprehensive and flexible pathway-based model.

Pathways should not be regarded as isolated individuals, complex inter-pathway dependences exist among pathways [19, 20], e.g., activation of the antigen processing and presentation pathway will lead to activation of other immune pathways. In this work, we constructed a pathway-pathway interaction (WWI) network which can describe the complex pathway-pathway dependencies to explore the drug synergy from a pathway-based perspective.

Since the topological relations between drug targets on the PPI network has been demonstrated to be able to estimate drug synergy to some extent, we ask if the pathway relations on the WWI network also possess a similar function or even prevail over target relations in determining whether the drug combination is synergistic or not. In order to test this hypothesis, we constructed a WWI network and comprehensively calculated the pathway-associations of different drugs based on the WWI network, thus evaluating whether synergistic drug combinations are more likely to target on inter-dependent pathways than random combinations.

According to a series of recent publications [21-25] in compliance with the 5-step rule [26], to develop an effective prediction method for a biomedical problem, we need to follow the five guidelines: (a) collect valid benchmark datasets to train and test the model; (b) formulate the samples by an effective feature value that can truly describe their intrinsic properties concerning the discussed problem; (c) develop a powerful method to implement the prediction; (d) objectively assess the performance of the proposed method; (e) establish a user-friendly web-server for the proposed method. Below, we try to predict drug synergy based on the WWI network following the above 5-step rule.

Results and discussion

Construction of a WWI network

A WWI network consisting of 269 nodes and 5991 edges was constructed as the basis of this study. Pathway information including the biological components, as well as pathway dependencies for each pathway was collected

from the KEGG [27] Markup Language (KGML) files, an exchange format of KEGG pathway maps. As of April 2015, there are 269 pathways which contain 6527 genes in total. To obtain the edges of WWI network, three kinds of pathway interactions were considered in the WWI network. First, manually annotated pathway dependencies, termed as functional-related WWIs, were extracted from the KGML files: if one pathway is recorded as one related pathway of another pathway in the KGML file, these two pathways can be taken as one functional-related WWI (see Materials and methods). In addition to these recorded relations, there must be other latent WWIs which are not curated in the KGML files at present. Intuitively, if two pathways are composed of shared biological components, there may be interactions between them- influence on one of them can be passed on to the others through the shared part. We calculated the significance (by P-value of Fisher's exact test) of gene-overlapping between two pathways, and pathway pairs with P-values less than a specific threshold 0.001, see Materials and methods for threshold determination were taken as gene-overlapping WWIS, i.e., the second kind of WWIs. In addition to the above two kinds of WWIs, pathways can also interact with each other by protein-protein interaction[20]. This type of interactions termed as protein-interacting WWIs are obtained by exploring pathway pairs with interacting coefficients (see Materials and methods) larger than certain threshold (0.001, see Materials and methods for the threshold determination). There are respectively 1100, 4929 and 12436 WWIs with respect to the functional-related, gene-overlapping and protein-interacting WWIs, where

4898 WWIs were redundant ones belonging to more than one type of WWIs (see Table S1). Most of the recorded pathway dependencies (860/1100) can also be recognized as gene-overlapping or protein-interacting WWIs. Meanwhile, most of the gene-overlapping WWIs (4038/4929) also interact with each other by PPIs. Besides, isolating pathways which are not connected with any pathway in the WWI network were removed from the WWIN.

The WWIN can provide a comprehensive description about pathway-pathway interactions. Taken the "NF-kappa B signaling pathway" (path:hsa04064) as one example, it is connected with 173 different pathways (see Table S2 for details), like "Toll-like receptor signaling pathway" (path:hsa04620), "TNF Signaling pathway" (path:hsa04668), and the pathway of "Hepatitis C" (path:hsa05160) on the WWI network. These interactions can cover the functional relations of different pathways. For example, Toll-like signaling pathway, which acts as primary sensors that detect microbial infections and induce innate immune responses can culminate in the activation of NF-kappaB [28]; TNF-alpha induces both survival and apoptotic signals, while the survival signal is mediated by the activation of NF-kappaB [29]; while the core protein of Hepatitis C virus can potentiate the activation of NF-kappaB [30]. Therefore, we make an assumption based on these functional associations among pathways, if a drug can act on one pathway, it may exert indirect impact on the other neighbor pathways which may be functional-related with the targeted one, and vice versa. With extensive description about pathway relations, the WWI network can also provide an opportunity to discover the

pathway associations between drugs and illustrate how different drugs cooperative with each other by acting on correlated pathways.

WWI network-based drug synergy evaluation

Based on the hypotheses that synergistic drugs may target on inter-dependent pathways [9, 15, 31], we defined a WWI network-based synergy score (WNS-score) to evaluate and prioritize drug pairs considering the topological associations of pathways on the WWI network. Two types of topological associations were considered here. The first one corresponds to the connectivity of different pathways on the WWI network. It is straightforward to calculate the connectivity simply based on the shortest path length between pathways if different drugs can act on totally different pathways. However, two drugs may exert influences on the same pathway, termed as co-regulated pathway in this study, through different or same targets, at the meantime. Under such condition, we take another type of topological association, the connectivity of targets on the PPI network, instead, to evaluate the synergistic association generated by the co-regulated pathway. Given two drugs A and B, we first collected their targets and projected them to the corresponding pathways. Next, the WNS-score of A and B was calculated based on the distances between pathways targeted by A and B on the WWI network, where the distance with respect to the co-regulated pathway is transformed from the distance of targets on the PPI-network (Figure 1, Materials and Methods). Thus, WNS-score can reflect the pathway-oriented associations between different drugs, which not only depends on the inter-relations of different pathways but also on the

inner-connection of targets belonging to co-targeted pathways. All candidate drug pairs were ranked by the WNS-scores, the top-ranked ones are more likely to be synergistic combinations.

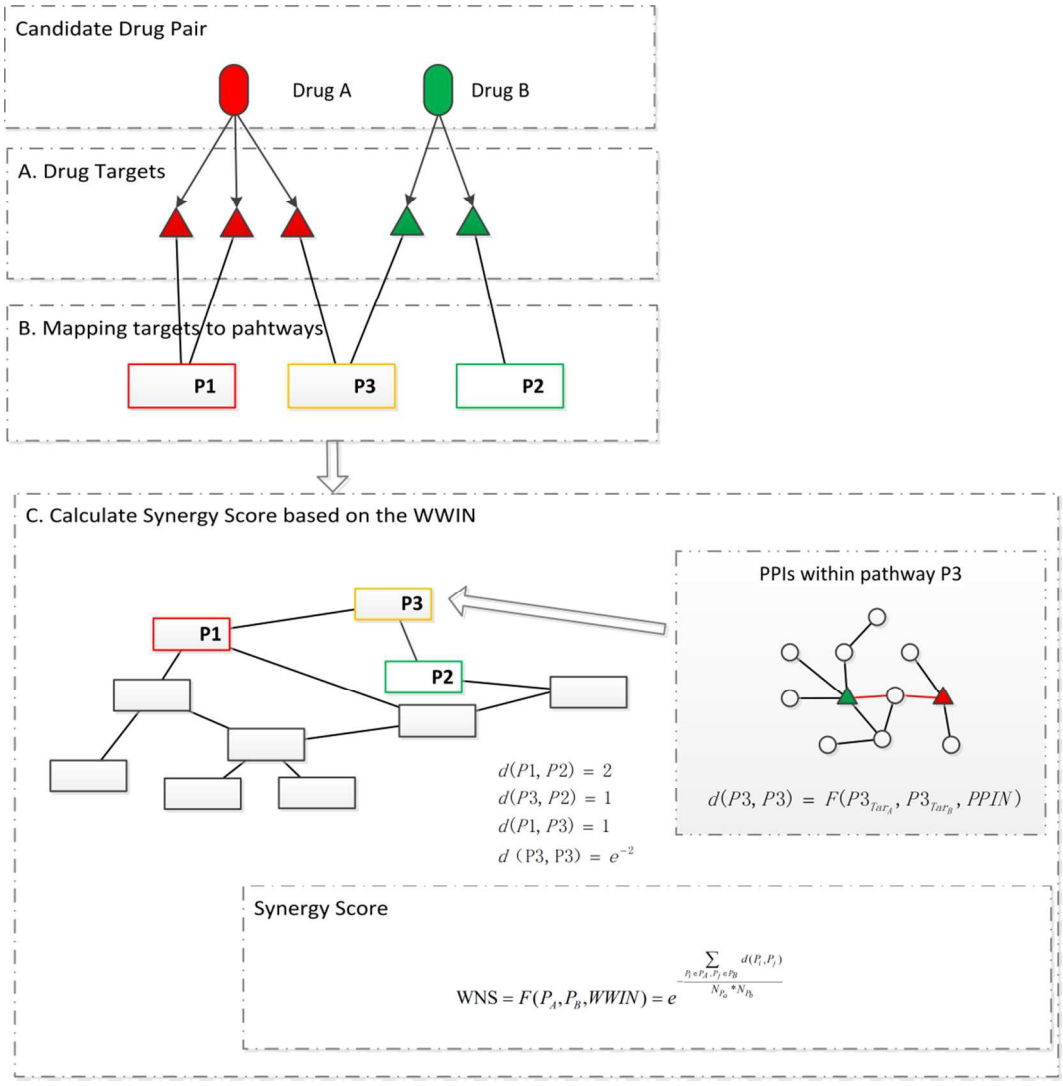


Figure 1. WWI network-based synergy. A. Collecting targets for each drug in a candidate drug combination. B. Mapping targets to pathways. C. Calculating synergy score based on the distances of all connected pair-wise pathways between drug A and B on the WWI network. When two drugs target on the

same pathways, like P3, the pathway distance in terms of this pathway, i.e. d(P3, P3), was transformed from the target connection on the PPI network.

Performance evaluation and comparison with other methods

To evaluate our synergy scoring scheme, we calculated the WNS-scores for both synergistic drug pairs (SDPs) and non-synergistic drug pairs (NSDPs) (see Table S3-S5 for details). The SDPs with confirmed synergistic effects were collected from the DCDB database version 2[32], while the NSDPs, which are utilized as negative contrasts to SDPs, were collected from DCDB and DrugBank[33] (see Materials and methods).

Three different kinds of pathway interactions including functional-related WWIs, protein-interacting WWIs and gene-overlapping WWIs were considered in this study (see Materials and methods). To see which kind of WWIs can contribute most to the effective prediction of SDPs, we tried to evaluate the performance of the WNS-scores based on different types of WWIs. We applied receiver operating characteristic (ROC) curves to estimate the performance of WNS-score in predicting SDPs. It turns out that the gene-overlapping WWIs can lead to the best prediction accuracy on all negative samples, and next to the gene-overlapping WWI, the protein-interacting ones performed better than functional-related WWIs (Figure 2). When three types of WWIs were merged together, the performance will be worse than that based on gene-overlapping or protein-interacting WWIs. This implies that pathway interactions predicted based on the shared genes may contribute most to predict drug synergy. The

shortage of knowledge on functional dependences among pathways, and the excessive amount of protein-interacting relations between pathways may underestimate or over-evaluate the interactions among pathways, thus reducing the performance of WNS-scores. We utilized the WNS-scores based on gene-overlapping WWIs to predict drug synergy in the following study.

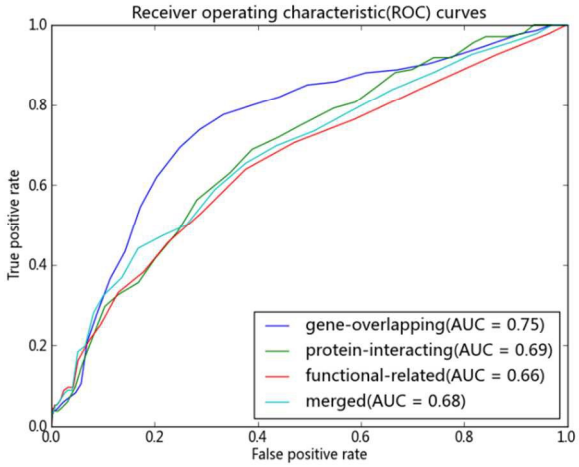


Figure 2. Comparison of the performances of WNS-scores based on different types of WWIs. The WNS-scores were respectively calculated based on the WWIN constructed by each single type of WWIs (gene-overlapping, protein-interacting and functional-related WWIs) and all types of WWIs merged together (denoted as "merged").

Next, we adopted fold enrichment analysis to test whether SDPs are with higher WNS-scores compared with NSDPs in the negative dataset. Results show top-ranked drug pairs are enriched by SDPs, and the number of SDPs decreased with the WNS-scores (Figure 3), i.e., synergistic drugs are more likely to get high WNS-scores compared to other combinations. According to the definition of WNS-score, it will be high when the number of connected

pathway pairs is small (but not zero) and the distances between pathways are short. It implies that two drugs are more likely to generate synergistic combination effects if these two drugs act on the same pathway through different targets or they cooperatively regulate a few of highly-associated pathways.

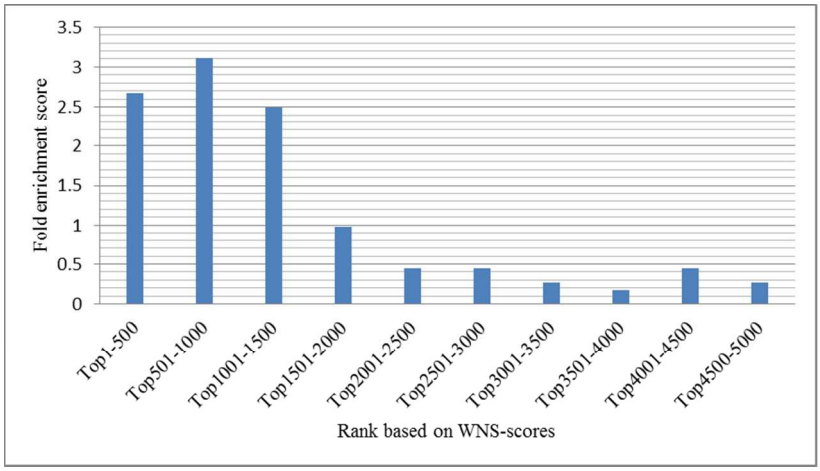


Figure 3. Fold-enrichment of SDPs. All drug pairs' WNS-scores were ranked from high to low. The x-axis represents the bin of rank, with each bin contain 500 drug pairs, while the y-axis represents the fold enrichment score of SDPs with in each bin.

To further verify the ability of WNS-score in predicting drug combination, we compared it with target-based prioritizing methods and other simple forms of pathway-based measurements: (1) target-overlapping (TarOverlap), calculating the target association by cosine similarity of two target sets, it is based on the hypothesis that drugs target on the same protein may generate synergistic effect [34]; (2) target distance (TarDis), calculating the target association based on the shortest path length between targets on PPI network [9];

(3) network-based multi-target synergy (NIMS), a method calculating the target associations considering both shortest path length between targets and node centrality on the PPI network [8]; (4) pathway-overlapping (PathOverlap), calculating the pathway association by cosine similarity of two pathway sets; (5) pathway distance (PathDis), calculating the pathway association based on the shortest path length between pathways on the WWIN; (6) a simplified version of WWI network-based synergy (SWNS) score, simplifying the WNS-score by overlooking the target association within co-regulated pathways and taking the distance between two same pathways as zero. What's more, to check whether the WNS-scores always perform better than other methods on different negative datasets, we evaluate the performances of these methods considering different negative datasets (see Materials and methods).

It turns out that all pathway-based measurements (pathOverlap, PathDis, WNS, SWNS) show better performance than target-based methods (TarOverlap, TarDis, NIMS) on all negative datasets, and our method, WNS-score, can achieve the best performance (Figure 4). Among all three negative datasets, the WNS-scores perform best (AUC: 0.82) when random drug pairs from DrugBank were taken as the negative dataset. When random drug pairs or unsuccessful drug pairs from DCDB were utilized as the negative dataset, the performance is decreased to 0.68 and 0.66. The reason may lie in that both datasets "NS2" and "NS3" are based on DCDB which aims at collecting synergistic drug combinations. These NSDPs based on drugs from DCDB are more likely to be similar with those SDPs which are also collected from DCDB

than those from DrugBank, thus making it difficult to distinguish SDPs and NSDPs. Although the performance varies when different negative datasets were adopted, the WNS-scores always out-perform other methods, in spite of which negative dataset was used. These comparisons prove that pathway associations between different drugs play an even more important role in estimating drug synergy. Besides, the simplified version of WNS-score, i.e., SWNS-score (see Materials and methods), although achieve better performance than most methods, is still inferior to WNS-score. This indicates that inclusion of the target associations within specific co-regulated pathways can help enhance the accuracy in predicting drug synergy. It also implies that target relations within specific pathway, like the upstream-downstream relations between targets of different drugs[35], also play key roles in generating synergistic effects. This WWI network-based synergy evaluation method which considers both the interactions between different pathways and the target associations within co-regulated pathways can provide one effective way to predict potential synergistic drug combinations.

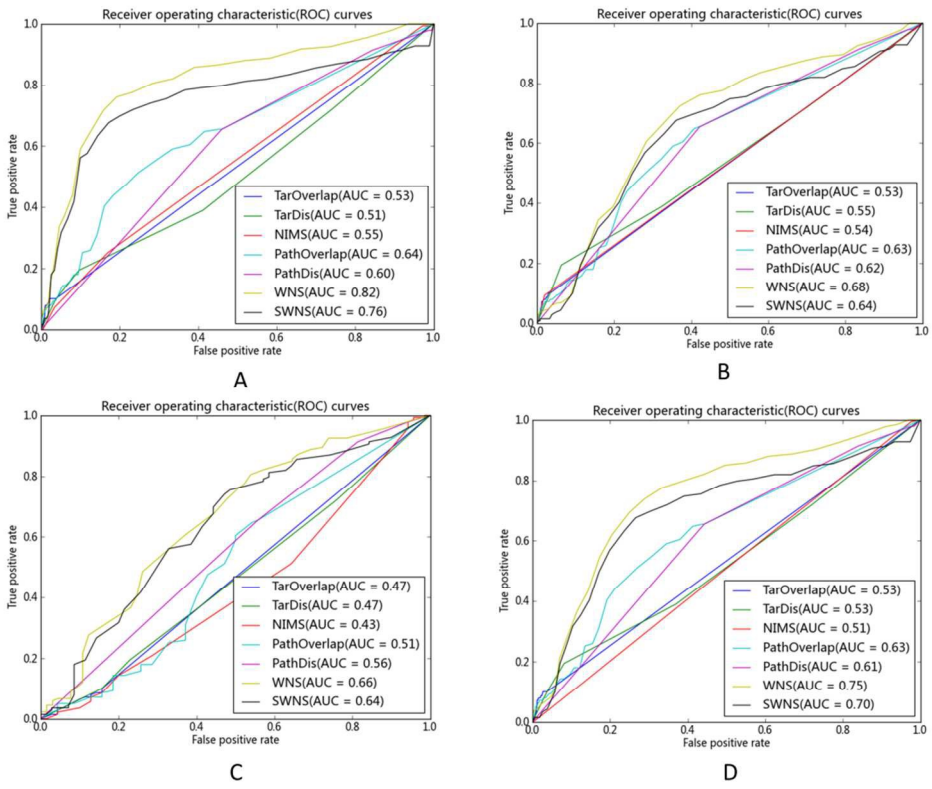


Figure 4. ROC curves to evaluate the performance of different methods based on different negative datasets. A. Performance evaluation based on random drug pairs from DrugBank, i.e., the negative dataset "NS1". B. Performance evaluation based on random drug pairs generated from DCDB, i.e., the negative dataset "NS2". C. Performance evaluation based on unsuccessful drug pair combinations in DCDB, i.e., the negative dataset "NS3". D. Performance evaluation based on all negative samples, i.e., all drugs pairs in the above three negative datasets.

Case studies on synergistic drug combinations

To further illuminate the superiority of WNS-score in predicting and explaining drug synergy, we adopted the combination of Gabapentin and Carbamazepine

as one example to illustrate how this WWI network-based method can help reveal the potential synergy mechanism. This combination is predicted with high WNS-score (0.9556) but low target-based scores (tarDis=0.049787, tarOverlap=0, NIMS=0.07778). Gabapentin and carbamazepine are both antiepileptic drugs. A preclinical isobolography study has proven that this combination can produce synergistic effects against epilepsy [36]. However, the latent mechanism is still unclear. According to the WWI network, both gabapentin and carbamazepine may both have effects on the pathway of "Adrenergic signaling in cardiomyocytes" (path:hsa04261) where the antiepileptic effects can be mediated through adrenergic alpha 1 or alpha 2 as indicated in [37, 38]. This co-regulation effect on "Adrenergic signaling in cardiomyocytes" through different targets may be the main reason for the synergistic outcome against epileptic.

Different from the above drug combination which only has one co-regulated pathway, more drug combinations may simultaneously act on more than one co-regulated pathways and inter-connected pathways. Taken another combination of naphazoline hydrochloride and pheniramine as an example, the scores for this combination were: WNS-score= 0.7165, tarOverlap=0, tarDis=0, and NIMS=0. This combination has been approved by FDA on the treatment of temporary relief of the minor eye symptoms of itching and redness. According to the WNS evaluation, it turns out that two pathways, "Neuroactive ligand-receptor interaction" (path:hsa04080) and "Calcium signaling pathway" (path:hsa04020), are co-regulated by these drugs through different targets. In

addition to the co-targeted pathways, two other pathways only targeted by Naphazoline hydrochloride can be related with the co-regulated pathways through one or two step network connection. Both the co-regulated and functional-related pathways may contribute to the synergistic effects of the drug combination.

Without remarkable target associations, neither of the above cases will be predicted as synergistic combinations by previous target-based approaches (tarDis, tarOverlap or NIMS). However, this novel WWI network-based method can distinguish them from random combinations and discover the potential synergy mechanism by exploring the WWIs between different drugs. As the above two cases indicate, drugs acting on the same pathway through different targets or drugs regulating a relatively small number of closely-correlated pathways are more likely to produce synergistic effects when combined.

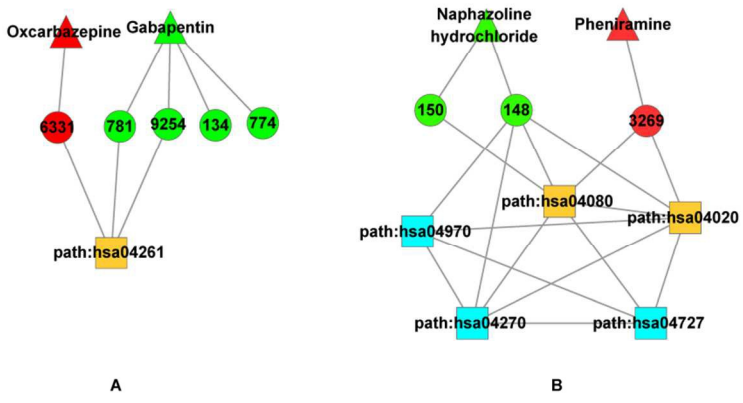


Figure 5. WWI-based synergy mechanism. A. Co-regulated pathways of oxcarbazepine and gabapentin. B. Co-regulated as well as inter-connected pathways

of naphazoline hydrochloride and pheniramine. Triangle, circle and rectangle nodes respectively represent drugs, targets, and pathways. The targets are labelled by gene Entrez ID [39], while the pathways are labelled by KEGG ID. To be noted, for clarity, not all targeted pathways for each drug are shown, and only those involved in the WNS-score calculation are presented.

Discussions

Synergistic drug combination has become a new promising way to improve the therapeutic management of complex diseases. With the development of computational methods in predicting the interactions between different biological elements[25, 40-43], more and more computational methods have been applied to investigate biomedical problems, drug synergy evaluation[8-12] is one of them. In this study we put forward for the first time a WWI network-based synergy evaluation method. This method can not only give rise to a much better performance than previous comparable target-based methods, but also provide an effective way to understand the potential synergy mechanism for some combinations. Compared with those methods incorporating various chemical information of drug combination[11, 44, 45], or models constructed on deliberate classification algorithm[10], this WNS-score presented in this study is much simpler and free of model training, but can still achieve a competitive performance. All these demonstrate the importance of pathway-pathway interactions in estimating drug synergy. Although obtained from in-silico studies, the WWI network-based findings have the potential to capture valuable insights and hypotheses on drug synergism.

However, limitations still exist for this method. First, this method is limited by the incomplete knowledge on pathway-pathway interactions as well as the pathway compositions, this WNS-score only works for drugs targeting on known pathways. Meanwhile, due to the incompleteness of the specific interaction effect like the potentiating or attenuating effect between pathways, we can only consider whether two pathways can interaction with each other or not, the specific interaction effect was still not taken into account. With more knowledge on the interaction effects, we will re-define the WNS-score based on a directed or weighted WWI network which can describe the interactions between pathways from a more comprehensive perspective. Second, although with relatively high prediction precision, only focusing on the interactions among pharmacological pathways which are interfered by drug targets is not enough, the synergy mechanism for some combinations may lie in the interactions between drug metabolizing processes. Therefore, more information about drug enzymes and corresponding metabolizing pathways should be incorporated in the future study. Besides, as suggested by some recent studies of classification or prediction methods[41-43], user-friendly and publically accessible web-servers will help improve the application values and influences [46], we shall make great efforts to provide a web-server for this WWI-based drug synergy evaluation method in our future work. With these further improvements, we believe that our method can be a promising way to supply guidance on drug combination development and provide meaningful hypothesis

on the synergy mechanism for further investigation. These mechanisms could be valuable for rational drug combination design among existing drugs.

Materials and methods

Data collection and preparation

Basic requirement for collected drugs. Since this study focus on the pathway correlations between different drugs, all drug pairs collected in the study must meet the premise that both drugs can be assigned with at least one pathway, i.e., we can obtain the target information for each drug and the targets can be projected to at least one pathway. The targets for drugs were collected from DrugBank and DCDB, and the target-pathway relationships were obtained from KEGG.

Positive data set. The positive data set was composed of synergistic drug pairs (SDPs). We collected drug combinations from the DCDB database, which collects and organizes information on drug combinations from clinical studies or the FDA orange book. Although there are 1363 drug combinations for the current version 2 of DCDB, a great part of them (1033/1363) is still investigational, and 237 of them were unsuccessful. To guarantee the effectiveness of selected SDPs, only 211 drug pairs which have been approved by FDA or recorded with synergistic effects in the DCDB records were taken into consideration. Further, according to the above basic requirement for drug collection, we retained 139 drug pairs from these drug combinations as the SDPs.

Negative datasets. Three different negative datasets composed of NSDPs were prepared. The first one, denoted as "NS1", was based on drugs of DrugBank. We randomly produced 5000 drug pairs which do not belong to drug combinations

recorded in DCDB. Further investigation on the availability of targets and pathways of each drug left 2948 drug pairs which meet the basic requirement. The second one ("NS2") was based on drugs from DCDB. The ones which meet the basic requirement for drug collection were taken as the candidate drugs, and we randomly produced additional 2948 pairs which are different from the recorded drug combinations of DCDB for dataset "NS2". To be more rigorous than random pairs, we also considered drug pairs recorded as unsuccessful combinations in the DCDB. Among them, 70 drug pairs which can meet the basic requirement were taken as the third negative data set ("NS3").

Pathway information collection. We downloaded all the KEGG xml (called KGML) files for pathways belonging to "Homo sapiens" from the KEGG pathway database (updated on April, 27, 2015). In the KGML file for one specific pathway, we can acquire both its gene members and pathway dependencies. Taken the pathway "Influenza A" as one example, one gene member for the pathway is recorded as an "entry" of type "gene":

<entry id="9" name="hsa:23586" type="gene"

link="http://www.kegg.jp/dbget-bin/www_bget?hsa:23586">

<graphics name="DDX58, RIG-I, RIGI, RLR-1" fgcolor="#000000" bgcolor="#BFFFBF"

type="rectangle" x="403" y="570" width="46" height="17"/>

</entry>

And one dependent pathway, like "Apoptosis" was recorded as one "entry" of type "map" as below:

<entry id="79" name="path:hsa04210" type="map"

link="http://www.kegg.jp/dbget-bin/www_bget?hsa04210">

<graphics name="Apoptosis" fgcolor="#000000" bgcolor="#FFFFFF"

type="roundrectangle" x="301" y="1037" width="76"

height="25"/>

</entry>

Parsing the KGML files, we can obtain extensive meaningful information which will assist in the construction of WWI network**.**

Network construction

PPI network. A PPI network was constructed based on all the PPIs from the HPRD database[47].

WWI network. A WWI network was constructed based on the pathway-pathway interactions (WWIs) among all KEGG pathways. In the WWI network, nodes are the "Homo sapiens" pathways, edges are based on the WWIs. Three types of WWIs were taken into consideration here. The first one, termed as functional-related WWIs, refers to the pathway dependencies recorded in the KGML files. For two pathways *P* and *Q*, if pathway *Q* is recorded as one "entry" of type "map" in the KGML file of *P*, or vice versa, these two pathways will be taken as one functional-related WWI. The second one is based on the gene-overlapping between different pathways[20]. If the number of shared genes between two pathways is significant against occasional condition (by Fisher's exact test, p-values of WWIs should be less than certain threshold), they are regarded as one gene-overlapping WWI. In addition, two pathways can also interact with each other by protein-protein interactions[20]. This type of interaction is termed as

protein-interacting WWI. The association was evaluated by the fraction of PPIs among two pathways. For example, if the number of proteins belonging to pathways *P* and *Q* are respectively *m* and *n*, and there are actually k pairs of PPIs between proteins of different pathways except of the overlapping part of two pathways, then an interacting coefficient was calculated as k/(m*n). If the coefficient is larger than certain threshold, there is one protein-interacting WWI between *P* and *Q*.**Network visualization.** All visualized networks were generated by Cytoscape[48], an open source software for visualizing biological networks.

WNS-score calculation

According to the synergy mechanism revealed by previous studies[9, 15, 49, 50], drugs with different but highly associated targets which belong to the same or related pathways are more likely to implement synergistic effects. Accordingly, a synergy score for each pair of compound ingredients was calculated based on the connectivity of pathways regulated by different drugs on the WWI network, as well as the inner-pathway's target associations for pathways regulated by both drugs.

The WWI network-based synergy score (*WNS-score*) for drugs D_A and D_B was calculated as:

$$WNS = F(P_A, P_B, G_{WWI}) = e^{-\frac{\sum_{P_i \in P_A, P_j \in P_B, linked(P_i, P_j)=1} d(P_i, P_j)}{N_{P_A} N_{P_B}}}$$

where P_A and P_B are respectively the pathway sets affected by drug D_A and D_B, while P_i *and* P_j are respectively one of them; *linked*(P_i, P_j)=*1* means that only pathway pairs which can be linked on the WWI network are considered in the

calculation; and N_{PA}, N_{PB} are respectively the number of pathways affected by drug D_A and D_B. In this equation, if P_i and P_j are different pathways, the distance between them is calculated by the shortest path length; if P_i and P_j is the same pathway, termed as co-regulated pathway, the distance in terms of this pathway is transformed from the target connections on the PPI network:

$$d(P_i,P_j)=\begin{cases} SPL(P_i,P_j,G_{WWI}), & \text{if } P_i \neq P_j \\ F(T_{P_i,a},T_{P_i,b},G_{PPI})=e^{-\frac{\sum_{T_i \in T_{P_i,a},T_j \in T_{P_i,b},linked(T_i,T_j)=1} d(T_i,T_j)}{N_{T_{P_i,a}} N_{T_{P_i,a}}}}, & \text{if } P_i = P_j \end{cases}$$

where $T_{P_i,a}$ and $T_{P_i,b}$ respectively represent Drug A and Drug B's targets which belongs to the co-regulated pathway P_i, $SPL(P_i,P_j,G_{WWI})$ represents the shortest path length between P_i and P_j on the WWI network, and the target distance between two Targets T_i and T_j was defined as :

$$d(T_i,T_j)=\begin{cases} SPL(T_i,T_j,G_{PPI}), \text{ if } T_i \neq T_j \\ 0, \text{ if } T_i = T_j \end{cases}$$

where $SPL(T_i,T_j,G_{PPI})$ represents the shortest path length between T_i and T_j on the PPI network.

This WNS-score was applied to predict the potential synergistic relations between two drugs. All candidate drug pairs were ranked by this score, the top-ranked ones are more likely to be synergistic combinations.

For comparison purpose, a simplified version of WNS-score (SWNS-score) was also calculated. For SWNS-score, we didn't consider the target associations within certain co-targeted pathway, that is to say, when two drugs target on the same pathway, the pathway-distance between these two drugs was regarded as zero:

$$d'(P_i,P_j)=\begin{cases} SPL(P_i,P_j,G_{WWI}), & \text{if } P_i \neq P_j \\ 0, & \text{if } P_i = P_j \end{cases}$$

Except of the pathway distances concerning co-targeted pathways, the SWNS-score was calculated in the same way of WNS-score.

Performance evaluation

Fold enrichment analysis. To check whether the drug pairs with high WNS-scores are more likely to be synergistic combinations, all candidate drug pairs in both positive and negative datasets were ranked by the WNS-score and binned into groups of 500 drug pairs. A fold enrichment score is defined as $\frac{m/500}{M/N}$ [45], where m is the number of SDPs within one certain bin of 500 drug pairs, M is the number of all SDPs, and N is the number of all drug pairs investigated. The denominator represents the expected percentage of SDPs in any bin if all SDPs equally distribute in the ranked list, the nominator represents the observed percentage of SDPs in the bin. If this score is high for certain bin, it represents the SDPs are more likely to be ranked in that bin.

ROC curve. ROC curves were applied to evaluate the precision of different methods. A ROC curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold setting. TPR is defined as: $TPR = \frac{TP}{TP + FN}$, and FPR is defined as: $FPR = \frac{FP}{FP + TN}$, where TP, TN, FP and FN respectively represent the number of true positives, true negatives, false positives and false negatives. In this study, true positives refer to SDPs with scores larger than threshold; true negatives refer to NSDPs with scores less than threshold; false positives represent NSDPs with scores larger than

threshold; while SDPs with scores less than threshold are regarded as false negatives. The area under the ROC curve (AUC) can describe the ability of an evaluation score to rank synergistic drug pairs higher than random ones.

Thresholds determination

To determine the thresholds for both the p-values of gene-overlapping WWIs and the interacting coefficient of protein interaction WWIs, we tried to calculate the WNS-scores based on WWIs under different thresholds, and the one which can lead to the best performance in distinguishing SDPs and NSDPs was utilized as the final threshold. As shown in Figure 5, gene-overlapping WWIs with p-values less than 0.001 can lead to the best performance (AUC: 0.75), and protein-interacting WWIs with coefficients larger than 0.001 perform best (AUC: 0.69). Consequently, these best-performance thresholds were taken as the final thresholds in the study. There were respectively 4929 gene-overlapping WWIs and 12436 protein-interacting WWIs, with 4038 WWIs belonging to both types.

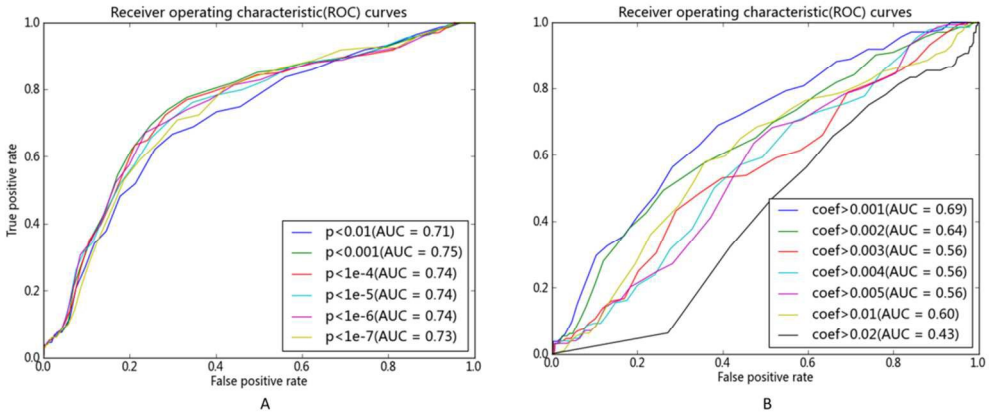


Figure 6. Performance of WNS-scores based on WWIs under different thresholds. The ROC curves were plotted according to the WNS-scores of the

positive and all three negative datasets. A. ROC curves of WNS-scores based on gene-overlapping WWIs determined by different thresholds of p-values. B. ROC curves of WNS-scores based on protein-interacting WWIs determined by different thresholds of interacting coefficient.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 81303152)

Supplementary information

Table S1. The WWI network. The first and second columns respectively represent two nodes on one edge, and the third column stands for the type of WWI.

Table S2. Neighbours of "NF-kappa B signaling pathway" in the WWI network.

Table S3. Scores for SDPs based on different methods.

Table S4. Scores for NSDPs in NS1 based on different methods

Table S5. Scores for NSDPs in NS2 based on different methods

Table S6. Scores for NSDPs in NS3 based on different methods

References

1. K. Traynor, *American journal of health-system pharmacy : AJHP : official journal of the American Society of Health-System Pharmacists*, 2012, **69**, 1616.
2. B. Al-Lazikani, U. Banerji and P. Workman, *Nature biotechnology*, 2012, **30**, 679-692.
3. T. C. Chou, *Cancer research*, 2010, **70**, 440-446.
4. G. R. Zimmermann, J. Lehar and C. T. Keith, *Drug discovery today*, 2007, **12**, 34-42.
5. S. Loewe, *Arzneimittel-Forschung*, 1953, **3**, 285-290.
6. R. J. Tallarida, *Genes & cancer*, 2011, **2**, 1003-1008.
7. T. C. Chou, *Pharmacological reviews*, 2006, **58**, 621-681.
8. S. Li, B. Zhang and N. Zhang, *BMC systems biology*, 2011, **5 Suppl 1**, S10.

9. Y. Y. Wang, K. J. Xu, J. Song and X. M. Zhao, *BMC bioinformatics*, 2012, **13 Suppl 7**, S7.
10. J. Zou, P. Ji, Y. L. Zhao, L. L. Li, Y. Q. Wei, Y. Z. Chen and S. Y. Yang, *Molecular bioSystems*, 2012, **8**, 3185-3196.
11. X. M. Zhao, M. Iskar, G. Zeller, M. Kuhn, V. van Noort and P. Bork, *PLoS computational biology*, 2011, **7**, e1002323.
12. L. Chen, B. Q. Li, M. Y. Zheng, J. Zhang, K. Y. Feng and Y. D. Cai, *BioMed research international*, 2013, **2013**, 723780.
13. A. M, G. VL, G. D, C. M, T. A, N. DJ and W. MJ, *oncotarget*, 2013, **4**, 14.
14. P. Dent, D. T. Curiel, P. B. Fisher and S. Grant, *Drug Resistance Updates*, 2009, **12**, 65-73.
15. J. Jia, F. Zhu, X. Ma, Z. Cao, Y. Li and Y. Z. Chen, *Nature reviews. Drug discovery*, 2009, **8**, 111-128.
16. K. A. Janes and D. A. Lauffenburger, *Journal of cell science*, 2013, **126**, 1913-1921.
17. B. Klinger, A. Sieber, R. Fritsche-Guenther, F. Witzel, L. Berry, D. Schumacher, Y. Yan, P. Durek, M. Merchant, R. Schafer, C. Sers and N. Bluthgen, *Molecular systems biology*, 2013, **9**, 673.
18. S. Nelander, W. Wang, B. Nilsson, Q. B. She, C. Pratilas, N. Rosen, P. Gennemark and C. Sander, *Molecular systems biology*, 2008, **4**, 216.
19. C. L. Hsu and U. C. Yang, *Bmc Genomics*, 2012, **13**.
20. K. Q. Liu, Z. P. Liu, J. K. Hao, L. Chen and X. M. Zhao, *BMC bioinformatics*, 2012, **13**, 126.
21. W. Chen, P. M. Peng, E. Z. Deng, H. Lin and K. C. Chou, *Anal Biochem*, 2014, **462**, 76-83.
22. H. Lin, E. Z. Deng, H. Ding, W. Chen and K. C. Chou, *Nucleic acids research*, 2014, **42**, 12961-12972.
23. B. Liu, L. Y. Fang, F. L. Liu, X. L. Wang, J. J. Chen and K. C. Chou, *Plos One*, 2015, **10**.
24. Y. Xu, X. Wen, L. S. Wen, L. Y. Wu, N. Y. Deng and K. C. Chou, *Plos One*, 2014, **9**.
25. J. H. Jia, Z. Liu, X. Xiao, B. X. Liu and K. C. Chou, *Journal of theoretical biology*, 2015, **377**, 47-56.
26. K. C. Chou, *Journal of theoretical biology*, 2011, **273**, 236-247.
27. M. Kanehisa and S. Goto, *Nucleic Acids Res.*, 2000, **28**, 27-30.
28. T. Kawai and S. Akira, *Trends in Molecular Medicine*, 2007, **13**, p460-469.
29. S. Gupta, R. Bi, C. Kim, S. Chiplunkar, L. Yel and S. Gollapudi, *Cell Death Differ*, 2005, **12**, 177-183.
30. L. R. You, C. M. Chen and Y. H. Lee, *Journal of virology*, 1999, **73**, 1672-1681.
31. P. Dent, D. T. Curiel, P. B. Fisher and S. Grant, *Drug resistance updates : reviews and commentaries in antimicrobial and anticancer chemotherapy*, 2009, **12**, 65-73.
32. Y. Liu, Q. Wei, G. Yu, W. Gai, Y. Li and X. Chen, *Database : the journal of biological databases and curation*, 2014, **2014**, bau124.
33. D. S. Wishart, C. Knox, A. C. Guo, S. Shrivastava, M. Hassanali, P. Stothard, Z. Chang and J. Woolsey, *Nucleic acids research*, 2006, **34**, D668-672.
34. M. A. Yildirim, K. I. Goh, M. E. Cusick, A. L. Barabasi and M. Vidal, *Nature biotechnology*, 2007, **25**, 1119-1126.

35. A. Korkut, W. Q. Wang, E. Demir, B. A. Aksoy, X. H. Jing, E. J. Molinelli, O. Babur, D. L. Bemis, S. O. Sumer, D. B. Solit, C. A. Pratilas and C. Sander, *Elife*, 2015, **4**.
36. S. J. Czuczwar, J. Kaplanski, G. Swiderska-Dziewit, A. Gergont, S. Kroczka and M. Kacinski, *Expert Opin Drug Met*, 2009, **5**, 131-136.
37. J. Azure, S. Poitra, B. Nelson, B. Goldenstein, C. Jurgens, D. Weinshenker and V. Doze, *Faseb J*, 2010, **24**.
38. C. R. Chen, W. M. Qu, M. H. Qiu, X. H. Xu, M. H. Yao, Y. Urade and Z. L. Huang, *Neuropharmacology*, 2007, **53**, 534-541.
39. D. Maglott, J. Ostell, K. D. Pruitt and T. Tatusova, *Nucleic acids research*, 2005, **33**, D54-D58.
40. X. Xiao, J. L. Min, W. Z. Lin, Z. Liu, X. Cheng and K. C. Chou, *J Biomol Struct Dyn*, 2015, **33**, 2221-2233.
41. J. L. Min, X. Xiao and K. C. Chou, *BioMed research international*, 2013.
42. X. Xiao, J. L. Min, P. Wang and K. C. Chou, *Plos One*, 2013, **8**.
43. X. Xiao, J. L. Min, P. Wang and K. C. Chou, *Journal of theoretical biology*, 2013, **337**, 71-79.
44. P. Li, C. Huang, Y. X. Fu, J. A. Wang, Z. Y. Wu, J. L. Ru, C. L. Zheng, Z. H. Guo, X. T. Chen, W. Zhou, W. J. Zhang, Y. Li, J. X. Chen, A. P. Lu and Y. H. Wang, *Bioinformatics*, 2015, **31**, 2007-2016.
45. J. Huang, C. Niu, C. D. Green, L. Yang, H. Mei and J. D. Han, *PLoS computational biology*, 2013, **9**, e1002998.
46. K. C. Chou, *Med Chem*, 2015, **11**, 218-234.
47. T. S. K. Prasad, R. Goel, K. Kandasamy, S. Keerthikumar, S. Kumar, S. Mathivanan, D. Telikicherla, R. Raju, B. Shafreen, A. Venugopal, L. Balakrishnan, A. Marimuthu, S. Banerjee, D. S. Somanathan, A. Sebastian, S. Rani, S. Ray, C. J. H. Kishore, S. Kanth, M. Ahmed, M. K. Kashyap, R. Mohmood, Y. L. Ramachandra, V. Krishna, B. A. Rahiman, S. Mohan, P. Ranganathan, S. Ramabadran, R. Chaerkady and A. Pandey, *Nucleic Acids Res.*, 2009, **37**, D767-D772.
48. P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski and T. Ideker, *Genome Res.*, 2003, **13**, 2498-2504.
49. N. Yi, W. Z. Ma, J. F. Pei, Q. Ouyang, C. Tang and L. H. Lai, *PloS one*, 2014, **9**.
50. Z. Li, H. Zhou, Y. Lu and T. Colatsky, *CPT: pharmacometrics & systems pharmacology*, 2014, **3**, e135.