Exploring Drug Combinations in A Drug-Cocktail Network

Ke-Jia Xu, Fu-Yan Hu
Department of Mathematics,
Institute of Systems Biology,
Shanghai University,
Shanghai 200444, China

Jiangning Song
State Engineering Laboratory for
Industrial Enzymes, Tianjin Institute
of Industrial Biotechnology
Chinese Academy of Sciences,
Tianjin 300308, China

Xing-Ming Zhao*
Institute of Systems Biology,
Shanghai University,
Shanghai 200444, China,
*Corresponding author, Email:
zhaoxingming@gmail.com

Abstract—**Combination of different agents is widely used clinically to combat complex diseases with improved therapy and decreased side effects. It is necessary to understand the underlying mechanisms of drug combinations. In this work, we proposed a network-based approach to investigate drug combinations. Our results showed that the agents in an effective combination tend to have more similar therapeutic effects and more interaction partners in a 'drug-cocktail network' than random combination networks. Based on our results, we further developed a statistical model termed as Drug Combination Predictor (DCPred) by using the topological features of the drug-cocktail network, and assessed its prediction performance by making full use of a well-prepared dataset containing all known effective drug combinations extracted from the Drug Combination Database (DCDB). As a result, our model achieved the overall best AUC (Area Under the Curve) score of 0.92. Our findings provide useful insights into the underlying rules of effective drug combinations and offer important clues as to how to accelerate the discovery process of new combination drugs in the future.**

Keywords-drug combination; drug-cocktail network; ATC codes

I. INTRODUCTION

Drug combination is the combination of different agents that can achieve better efficacy but less side effects compared to its single component. Recently, it is becoming a popular and promising strategy to new drug discovery, especially for treating complex diseases, e.g. cancer.[1-3] For example, Moduretic is the combination of Amiloride and Hydrochlorothiazide, which is an approved combination used to treat patients with hypertension.[4, 5] Chan *et al.* [6] identified a combination drug, namely Tri-Luma, for combating melasma (dark skin patches) of the face based on efficacy and safety experiments. Despite of the efforts that have been made to discover new drug combinations in the past few decades, the majority of the effective combination drugs clinically used were discovered through experiences that generally required the labor-intensive and time-consuming "brute force" screening of all possible combinations of the approved individual drugs.[7] In a drug combination, a drug may promote or suppress the effect of another one. For instance, cyclosporine increases the effect of sirolimus, while bupropion decreases the effect of cyclosporine. As a result, two drugs may have a totally new effect that is different and not expected from either individual drug.[8, 9] Furthermore, the number of possible combinations will increase exponentially with the increasing availability of single drugs. For example, in the case of four drugs, there will be six possible combinations. This number would be enormous considering the fact that there are thousands of approved drugs. Due to the huge search space of possible combinations between known drugs, the identification of optimal and effective drug combinations is a non-trivial task.

Therefore, it is necessary to develop effective *in silico* methods that are capable of discovering new drug combinations prior to combination synthesis and practical test in the lab. Owing to the completion of human genome sequencing projects and the advancement of molecular medicine, extensive system biology efforts have been made to discover new combinations based on molecular interaction networks in the past few years.[10-13] In this context, Geva-Zatorsky *et al.* [10] have recently found that the protein dynamics in response to drug combination can be accurately described by a linear superposition of the dynamics under the corresponding individual drugs. Calzolari *et al.* [11] devised an efficient search algorithm originated from information theory for the optimization of drug combinations based on the sequential decoding algorithms. More recently, researchers have also developed computational frameworks for predicting drug combinations and synergistic effects based on high-throughput data. [12, 13]

In general, the binding of a drug to its target proteins will in turn affect the downstream pathways or biological processes pertinent to this drug.[14] In other words, drugs work by affecting biological systems that are perturbed by their targets. In the case of combination drugs, different agents that have different action mechanisms by targeting different biochemical pathways or molecular interaction networks are combined to control the multi-factor regulated processes of complex disease. Therefore, the network circuits where drug target proteins function should provide useful insight into the action mechanisms of the combination drugs.

978-1-4577-1666-9/11/$26.00 ©2011 IEEE

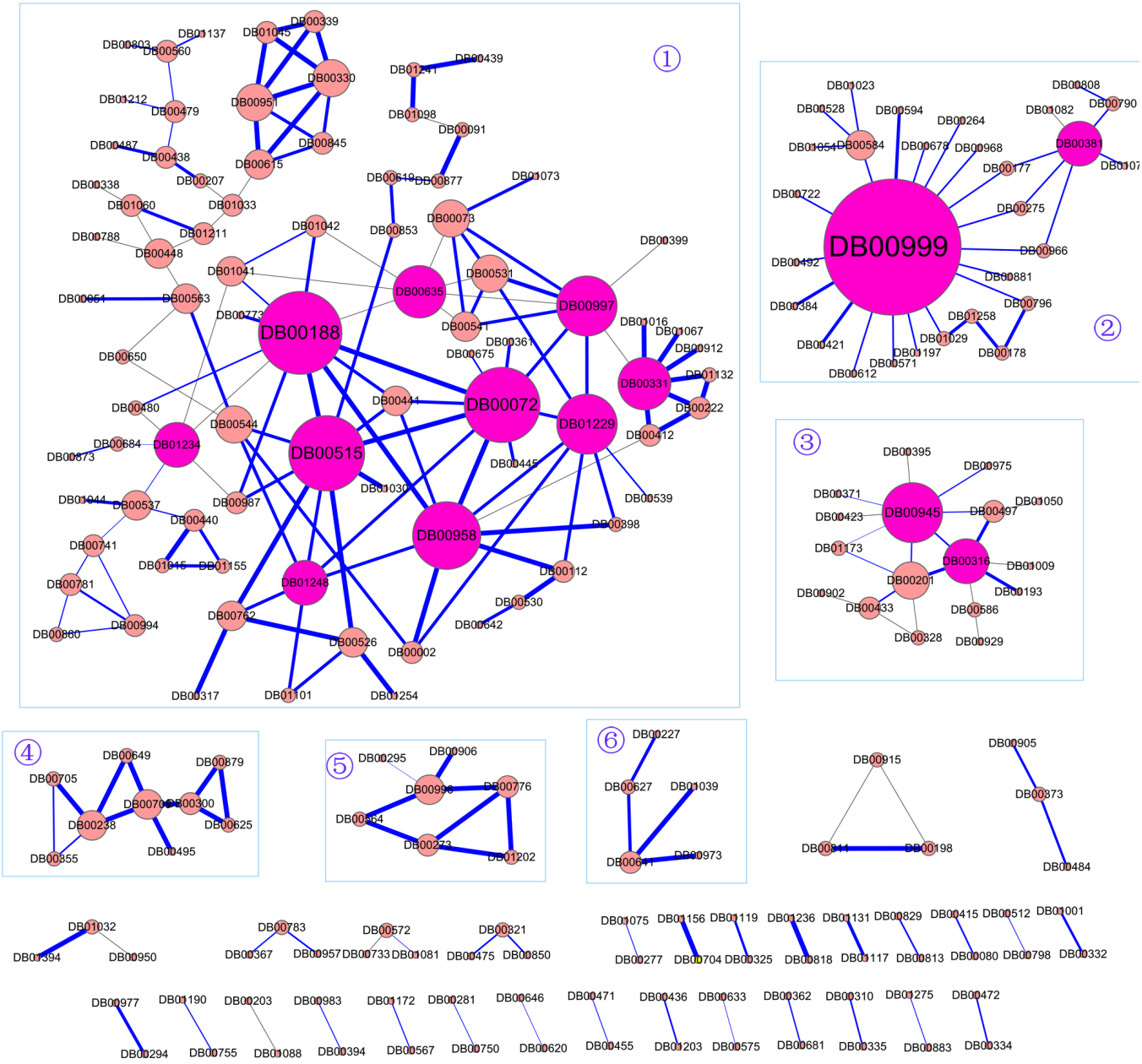

Figure 1. The 'drug-cocktail network'. The nodes represent drugs and an edge denotes an effective combination consisting of the two drugs linked by the edge. The hub drugs that have more than 6 neighbors are colored in purple. The size of each node approximates its degree, the width of each edge approximates the therapeutic similarity (see Equation 2) between the two drugs linked by the edge, and a grey edge means that the two drugs linked by that edge have totally different therapeutic effects. The numbers in panel 1-6 represent the top 6 largest connected components from the whole drug-cocktail network.

In our recent work, a drug-target network biology approach was developed to describe the underlying rules of drug combinations.[15] In particular, we found that the target proteins of effective drug combinations tended to be located in the close proximity of protein interaction networks and involved in functionally related pathways or biological processes.

In this work, we studied the combination drugs according to their therapeutic similarity and the network topology of the drug-cocktail networks constructed from the effective drug combinations in the Drug Combination Database (DCDB)[16]. We found that the drugs in an effective combination tend to have more similar therapeutic effects and more protein interaction partners in the form of drug-cocktail networks than random combination networks. We further developed a statistical model called DCPred to predict possible drug combinations and validated this model based on a benchmark dataset with the known effective drug combinations. At the end, DCPred model achieved the overall best AUC (Area Under the Curve) score of 0.92.

978-1-4577-1666-9/11/$26.00 ©2011 IEEE

II. METHODS

Data sources

The annotations of drug combinations were retrieved from a newly released Drug Combination Database (DCDB)[16]. This is a major resource for collecting effective drug combinations from the literature. The target protein information, the ATC code annotation of the drugs and protein subcellular localizations, were extracted from DrugBank[17]. Drug combinations that do not have ATC codes for the corresponding drug components and combinations with none or unclear efficacy were discarded. Consequently, 194 effective drug combinations were obtained, including 76 approved combinations, 64 clinical combinations and 54 preclinical combinations. We then split the combinations with more than two drug components into combination pairs, resulting in 239 drug combination pairs. They were used to construct a drug-cocktail network (Fig. 1), where the nodes represent drugs and the edges represent combinations, respectively. In the drug-cocktail network, the size of each node denotes its degree and the width of each edge denotes the therapeutic similarity (*TS*) between the two drugs linked by the edge. The gray edge means that there is no therapeutic similarity between the two drugs.

Drug therapeutic similarity

The Anatomical Therapeutic Chemical (ATC) Classification System, which includes five different hierarchical levels, was used to classify drugs into different groups according to the organ they acted on and the therapeutic chemical characteristics. The *k*-th level drug therapeutic similarity (S_k) between two drugs is defined using the ATC codes of these two drugs:

$$S_k(d_1, d_2) = \frac{ATC_k(d_1) \cap ATC_k(d_2)}{ATC_k(d_1) \cup ATC_k(d_2)} \quad (1)$$

where $ATC_k(d)$ denotes all the ATC codes at the *k*-th level of drug *d*. Note that a drug has five levels of ATC codes. A score, *TS*, is used to define the therapeutic similarity between two drugs:

$$TS(d_1, d_2) = \frac{\sum_{k=1}^{n} S_k(d_1, d_2)}{n} \quad (2)$$

where *n* ranges from 1 to 5. In this study, $n = 3$ is adopted considering that only very few drugs have same ATC codes at the 5th level.

Drug combination prediction model

We assume that two drugs are more likely to be combined if they share large numbers of common drugs in the drug-cocktail network. For example, if two drugs d_1 and d_2 with respective n_1 and n_2 partners have m in common in the drug-cocktail network, there will be three groups in the neighborhood of the two drugs, i.e. (1) m drugs that are the neighbors of both drug d_1 and d_2; (2) $n_1 - m$ partners that are the neighbors of drug d_1 only; and (3) $n_2 - m$ partners are the neighbors of drug d_2 only.[18] Suppose that there are totally N drugs in the drug combination network, then the probability for which d_1 and d_2 are combined can be calculated using the following equation

$$P(m, n_1, n_2, N) = \frac{\binom{N}{m}\binom{N-m}{n_1-m}\binom{N-n_1}{n_2-m}}{\binom{N}{n_1}\binom{N}{n_2}} \quad (3)$$

We built statistical models (termed as Drug Combination Predictor, DCPred) using the above equation (3) and assessed the performance for inferring effective drug combinations based on the curated drug combinations dataset.

III. RESULTS AND DISCUSSION

The drug-cocktail network

In this study, we extracted 239 known effective pairwise drug combinations from the Drug Combination Database (DCDB)[16]. The ATC code information for each drug was obtained from DrugBank[17]. Based on this dataset, we constructed the drug-cocktail network with 215 nodes and 239 edges (see Fig. 1 for the visualization of this network), where nodes represent the drugs and an edge is connected if the two drugs are found in an effective drug combination. In Fig.1, the size of each node approximates its degree, and the width of each edge approximates the therapeutic similarity (*TS*) (as defined in Equation 2) between the two drugs linked by the edge, while the grey edges indicate that the two drugs linked by the edge have totally different therapeutical effects. In total, 83.3% (199/239) of the combination pairs have therapeutic similarities. In addition, we found that 102 combination drugs had at least two neighbors in the drug-cocktail network, which we termed as "star drugs" hereafter and 91 of which had target protein annotations in DrugBank.

Since most of biological networks are scale-free networks [19], we analyzed the topology of the drug-cocktail network in order to find out whether it is also a scale-free network. The degree distribution of the drug-cocktail network is shown in Fig. 2. It is evident that the degree distribution follows a power law distribution, suggesting that it is indeed a scale-free network. That is, the fraction $P(x)$ of nodes in the drug-cocktail network having x connections to other nodes can be described as

$$p(x) \propto cx^{-\alpha} \quad (4)$$

where $c = 2.1$ and $\alpha = 1.9$, respectively.

978-1-4577-1666-9/11/$26.00 ©2011 IEEE

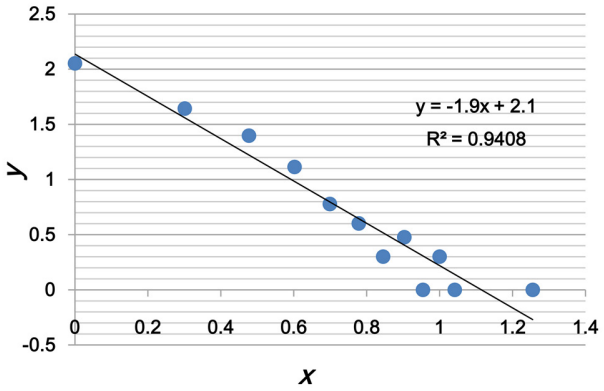

Figure 2. The degree distribution of the drug-cocktail network. The *x*-axis represents the common logarithm of the value of degree *k*, while the *y*-axis represents the common logarithm of the fraction of drugs that have the degree of *k*.

TABLE I. THE ENRICHED THERAPEUTIC EFFECTS REPRESENTED BY THE ATC CODES (FIRST LEVEL) FOR THE TOP 6 LARGEST CONNECTED COMPONENTS, WHERE THE NUMBERING FOR EACH CONNECTED COMPONENT IS CONSISTENT WITH THAT SHOWN IN FIG. 1. HERE, THE ENRICHED ATC CODES MEAN THAT THEY OCCUR MORE FREQUENTLY, EITHER MORE THAN 10 TIMES OR ACCOUNTING FOR MORE THAN 40% OF ALL ATC CODES ASSIGNED TO THE DRUGS IN THE CONNECTED COMPONENTS.

Connected component numbering	Number of drugs	Enriched ATC codes: Frequency
1	84	L:40, J:24, A:16, S:11
2	29	C:28
3	17	N:8, M:7
4	9	J:9
5	7	N:7
6	5	J:5

TABLE II. THE *P*-VALUE AT WHICH THE RATIO OF THE THERAPEUTIC SIMILARITY (*TS*) SCORE OF A RANDOM NETWORK IS LARGER THAN THAT OF THE DRUG-COCKTAIL NETWORK IN THE RANDOMIZATION TESTS OF 1000 TIMES AT DIFFERENT ATC CODE LEVELS.

ATC code level	1	2	3	4
P-value	0/1000	0/1000	0/1000	0/1000

As the drug-cocktail network shown in Fig. 1 was not fully connected, the top 6 largest connected components were chosen for further analysis. We will consider the drug-cocktail network as the union of these 6 connected components hereafter unless stated otherwise. In particular, each connected component was found to be enriched for one or several therapeutic classes according to the ATC classification system, as shown in Table I. In other words, the drugs having similar therapeutic effects tend to be clustered together in the drug-cocktail network.

To test our hypothesis that the combination drugs tend to have similar therapeutic effects, the drug-cocktail network was compared against random combination networks. For this purpose, a therapeutic similarity (*TS*) score was calculated for each drug combination pair, and the average of all *TS* scores was calculated as the *TS* score for the whole drug-cocktail network. The random combination networks were generated by randomly shuffling the edges while still preserving the degree for each node[20] in the drug-cocktail network, and this procedure was repeated for 1,000 times at different ATC code levels ranging from 1 to 4. To examine the statistical significance of the difference between the drug-cocktail network and random combination networks, the *P*-value was used and defined as the ratio that the *TS*s of random combination networks were larger than that of the drug-cocktail network among the 1000 randomizations. The results are shown in Table II. The calculated *P*-values of the drug-cocktail network across ATC code levels 1-4 are all

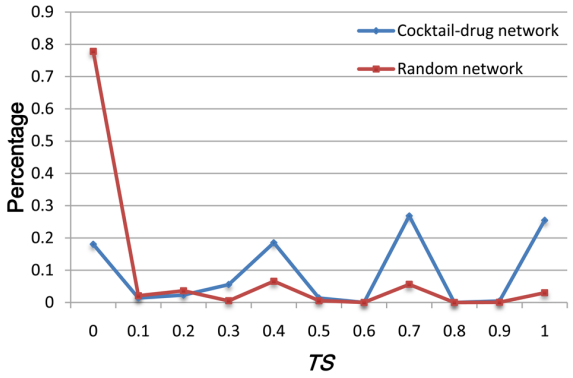

Figure 3. The distribution of the *TS* scores between star drugs and their neighbors. Blue and red lines represent the drug-cocktail network and random network, respectively.

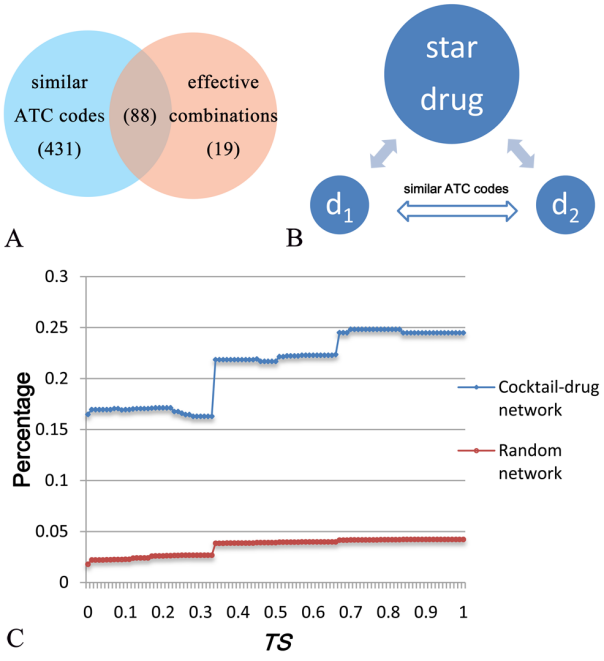

Figure 4. (A) The distribution of neighbor drug pairs of star drugs. The neighbor pairs of star drugs can be classified into two groups, according to whether they have similar ATC codes (the blue area), or whether they are used as effective combinations (the pink area). (B) The relationship between two neighbors d1 and d2 of a star drug. (C) The percentage of effective combinations within neighbor drug pairs with *TS* equal to or larger than a certain threshold. Blue and red lines represent the drug-cocktail network and the average of 1000 randomly generated combination networks, respectively

978-1-4577-1666-9/11/$26.00 ©2011 IEEE

equal to 0, strongly suggesting that the real drug combinations are different from the random combination networks. Note that the 5th ATC code level was not considered here, as there was only one drug combination having identical ATC codes for all 5 levels in the drug-cocktail network. This means that the 5th ATC code level was not suitable for performing statistical analysis and thus it was not included in the analysis.

Furthermore, we studied the therapeutic effects of the "star drugs" and their neighbors in the drug-cocktail network in order to reveal whether star drugs have therapeutic similarities to all their neighbors. Fig. 3 shows the distribution of the *TS* scores relative to star drugs and their neighbors. For the effective combination pairs involving star drugs, 82% have therapeutic similarity, and most of star drugs have similar therapeutic effects as the majority of their neighbors. As a contrast, 78% of the combination pairs in the random network do not have any therapeutic similarity. These results suggest that the star drugs tend to be used in combination with drugs that have therapeutic similarity.

Moreover, we also investigated the distribution of neighbor drug pairs of star drug (Fig. 4A and 4B), attempting to answer whether or not the drug combination pairs that share a star drug have therapeutic similarity. To do so, we divided the neighbor drug pairs of a star drug into two groups, according to whether they have similar ATC codes, or whether they are approved effective combinations. We then calculated the percentage of effective combinations among drug pairs that share a star drug and have a *TS* score equal to or larger than a certain threshold (Fig. 4C). From Fig. 4C, we can see that the more similar therapeutic effects (as reflected by the *TS* score) the two drugs have, the more likely they are effective combinations. Another observation is that the combinations between drugs sharing similar therapeutic effects and star drugs tend to be more effective combinations.

Implication of drug networks for possible drug combinations

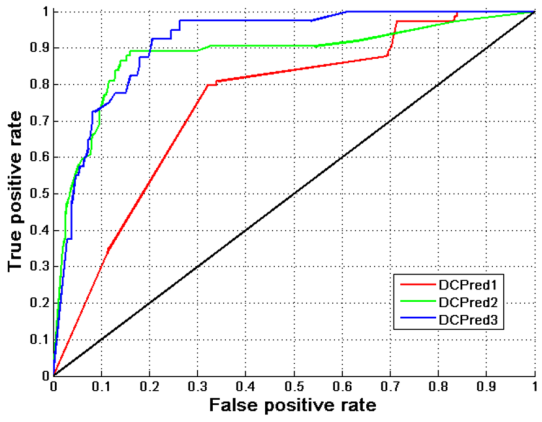

Figure 5. The ROC curves of different DCPred models, where DCPred1 uses *TS* only, DCPred2 uses *TS* and drugs with at least 2 neighbors, and DCPred3 uses *TS* and drugs with at least 3 neighbors.

As shown in Fig. 3, 82% of the combinations between star drugs and their neighbors have therapeutic similarity, and most of the star drugs have therapeutic similarity to the majority of their neighbors in the drug-cocktail network. Additionally, most of the effective combinations are observed to be located in the vicinity of drug pairs with similar ATC codes. Hence, it is possible to predict drug combinations from the set of drug pairs with similar ATC codes. However, we found that there are only 74 known effective combinations in all of the 1181 possible combinations with similar ATC codes. Since the number of effective drug combinations is considerably smaller than that of random combinations between drugs having similar ATC codes, it is a challenging but crucial task to discover the effective combinations from the pool with a vast number of random combinations.

In Fig. 4B and 4C, we can see that if two drugs with similar ATC codes have a common neighbor in the drug-cocktail network, it is more likely that they are combined together. Therefore, we assumed that the two drugs having similar ATC codes and sharing a significantly larger number of common partners in the drug-cocktail network are more likely to be combined effectively. Based on this assumption, we further developed a new statistical model called DCPred to test this hypothesis and applied it to predict and rank all the possible drug combinations (See Materials and methods for more details). All possible drug combinations were ranked, and the top ones were considered as putative effective drug combinations. The list of drug combination ranking can be found in the additional files. We found that two drugs with more common neighbors generally have higher rankings. Consequently, we got 74 effective drug combinations as the positive set and 1107 combinations without any annotations as the random set (Supplementary Table 1). To assess the prediction performance of each DCPred model, we plotted the ROC curves [21] in Fig. 5, where the drug pairs ranked above a given threshold were predicted as effective drug combinations (positive), while the rest were regarded as negative. As a comparison, we plotted the ROC curve of the model (DCPred1, red curve in Fig. 5) based on *TS* only, and compared its results to those ranked by our models (DCPred2 uses *TS* and drugs with at least 2 neighbors, and DCPred3 uses *TS* and drugs with at least 3 neighbors) for the same data set. We then calculated the area under the ROC curves (AUC)[22] for the different DCPred models. As a result, DCPred2 achieved an AUC score of 0.88 (the green curve in Fig. 5), in comparison with AUC of 0.75 for the *TS*-based method (DCPred1) (the red curve in Fig. 5).

If we only considered the combinations whose drug components have at least 3 neighbors, we obtained 40 positive samples and 379 random samples (Supplementary Table 2). Accordingly, the resulting model was termed as DCPred3 (the blue curve in Fig. 5). It achieved the AUC score of 0.92. Compared with the aforementioned two models DCPred1 and DCPred2, based on the information of at least 3 neighor drugs, DCPred3 model resulted in the overall best performance. We hope that the DCPred models developed in this study can be used to facilitate the *in silico* identification of effective drug combinations and speed up the future discovery process.

978-1-4577-1666-9/11/$26.00 ©2011 IEEE

IV. CONCLUSIONS

Drug combination is a promising strategy for combating complex disease, but our complete understanding of the underlying mechanisms of drug combinations is largely lacking at present. It is therefore necessary to develop efficient screening techniques to infer effective drug combinations in order to reduce the labor-intensive, time consuming trial-and-error experiments. In this article, we extracted all the known effective drug combinations from DCDB and then constructed a drug-cocktail network, including 215 drugs and 239 effective drug combinations. Based on this cocktail network, we observed that the star drugs tended to have therapeutic similarity with their drug neighbors, and two drugs having similar therapy and sharing neighbors tended to be employed in drug combination. Our analysis also revealed that: 1) hub drugs usually have similar and even the same therapeutic effects as their neighbors; 2) statistical test indicates that the components in effective drug combinations usually have more similar therapeutic effects and share more common neighbor drugs in the drug-cocktail network, making the drug-cocktail network differ from the random combination networks.

Based on the above observations, we developed a new statistical model to infer and rank possible effective drug combinations, taking into account drugs with at least two or three drug neighbors. As a result, our DCPred2 and DCPred3 models achieved the AUC scores of 0.88 and 0.92, respectively, demonstrating their high performance. Our results in this study provide useful insight into the underlying mechanism of effective drug combinations and hence offer useful clues for reducing the search space of possible combinations within the approved drugs. DCPred models are expected to be useful for developing more accurate models and can be applied to screen more effective drug combinations with clinical importance.

ACKNOWLEDGMENTS

XMZ was partly supported by the Innovation Program of Shanghai Municipal Education Commission (10YZ01, 09ZZ93), Innovation Funding of Shanghai University and Shanghai Rising-Star Program (10QA1402700). JS would like to thank the National Health and Medical Research Council of Australia (NHMRC) and the Chinese Academy of Sciences (CAS) for financially supporting this research via the NHMRC Peter Doherty Fellowship and the Hundred Talents Program of CAS. KJX was supported by the Innovation Program of Shanghai University (SHUCX112015).

REFERENCES

[1] A. Argiris, *et al.*, "Synergistic interactions between tamoxifen and trastuzumab (Herceptin)," *Clinical Cancer Research,* vol. 10, pp. 1409-1420, Feb 15 2004.
[2] C. K. Osborne and R. Schiff, "Growth factor receptor cross-talk with estrogen receptor as a mechanism for tamoxifen resistance in breast cancer," *Breast,* vol. 12, pp. 362-367, Dec 2003.
[3] J. C. Marsh, *et al.*, "The influence of drug interval on the effect of methotrexate and fluorouracil in the treatment of advanced colorectal cancer," *J Clin Oncol,* vol. 9, pp. 371-80, Mar 1991.
[4] D. R. Wilson, *et al.*, "Interaction of amiloride and hydrochlorothiazide with atrial natriuretic factor in the medullary collecting duct," *Can J Physiol Pharmacol,* vol. 66, pp. 648-54, May 1988.
[5] J. Frank, "Managing hypertension using combination therapy," *American Family Physician,* vol. 77, pp. 1279-1286, May 1 2008.
[6] R. Chan, *et al.*, "A randomized controlled trial of the efficacy and safety of a fixed triple combination (fluocinolone acetonide 0.01%, hydroquinone 4%, tretinoin 0.05%) compared with hydroquinone 4% cream in Asian patients with moderate to severe melasma," *Br J Dermatol,* vol. 159, pp. 697-703, Sep 2008.
[7] G. R. Zimmermann, *et al.*, "Multi-target therapeutics: when the whole is greater than the sum of the parts," *Drug Discovery Today,* vol. 12, pp. 34-42, Jan 2007.
[8] M. Pennati, *et al.*, "Potentiation of paclitaxel-induced apoptosis by the novel cyclin-dependent kinase inhibitor NU6140: a possible role for survivin down-regulation," *Molecular Cancer Therapeutics,* vol. 4, pp. 1328-1337, Sep 2005.
[9] B. R. Lewis, *et al.*, "Pharmacokinetic interactions between cyclosporine and bupropion or methylphenidate," *Journal of Child and Adolescent Psychopharmacology,* vol. 11, pp. 193-198, Sum 2001.
[10] N. Geva-Zatorsky, *et al.*, "Protein Dynamics in Drug Combinations: a Linear Superposition of Individual-Drug Responses," *Cell,* vol. 140, pp. 643-651, Mar 5 2010.
[11] D. Calzolari, *et al.*, "Search Algorithms as a Framework for the Optimization of Drug Combinations," *Plos Computational Biology,* vol. 4, p. e1000249, Dec 2008.
[12] G. Jin, *et al.*, "An enhanced Petri-net model to predict synergistic effects of pairwise drug combinations from gene microarray data," *Bioinformatics,* vol. 27, pp. i310-i316, Jul 1 2011.
[13] Z. Wu, *et al.*, "A systems biology approach to identify effective cocktail drugs," *BMC Syst Biol,* vol. 4 Suppl 2, p. S7, 2010.
[14] Q. L. Li and L. H. Lai, "Prediction of potential drug targets based on simple sequence properties," *Bmc Bioinformatics,* vol. 8, p. 353, Sep 20 2007.
[15] K. J. Xu, *et al.*, "A network biology approach to understand combination of drugs," *ISB2010,* pp. 347-354, 2010.
[16] Y. B. Liu, *et al.*, "DCDB: Drug combination database," *Bioinformatics,* vol. 26, pp. 587-588, Feb 15 2010.
[17] D. S. Wishart, *et al.*, "DrugBank: a knowledgebase for drugs, drug actions and drug targets," *Nucleic Acids Research,* vol. 36, pp. D901-D906, Jan 2008.
[18] M. P. Samanta and S. Liang, "Predicting protein functions from redundancies in large-scale protein interaction networks," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 100, pp. 12579-12583, Oct 28 2003.
[19] M. E. J. Newman, "The structure and function of complex networks," *Siam Review,* vol. 45, pp. 167-256, Jun 2003.
[20] S. Maslov and K. Sneppen, "Specificity and stability in topology of protein networks," *Science,* vol. 296, pp. 910-913, May 3 2002.
[21] D. Hand, "Measuring classifier performance: a coherent alternative to the area under the ROC curve," *Machine Learning,* vol. 77, pp. 103-123, 2009.
[22] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters,* vol. 27, pp. 861-874, Jun 2006.

978-1-4577-1666-9/11/$26.00 ©2011 IEEE