A simple gene set-based method accurately predicts the synergy of drug pairs

Yu-Ching Hsu†, Yu-Chiao Chiu†, Yi-Dong Chen‡, Tzu-Hung Hsiao§ and Eric Y. Chuang§

From The International Conference on Intelligent Biology and Medicine (ICIBM) 2015
Indianapolis, IN, USA. 13-15 November 2015

Abstract

Background: The advance in targeted therapy has greatly increased the effectiveness of clinical cancer therapy and reduced the cytotoxicity of treatments to normal cells. However, patients still suffer from cancer relapse due to the occurrence of drug resistance. It is of great need to explore potential combinatorial drug therapy since individual drug alone may not be sufficient to inhibit continuous activation of cancer-addicted genes or pathways. The DREAM challenge has confirmed the potentiality of computational methods for predicting synergistic drug combinations, while the prediction accuracy can be further improved.

Methods: Based on previous reports, we hypothesized the similarity in biological functions or genes perturbed by two drugs can determine their synergistic effects. To test the feasibility of the hypothesis, we proposed three scoring systems: co-gene score, co-GS score, and co-gene/GS score, measuring the similarities in genes with significant expressional changes, enriched gene sets, and significantly changed genes within an enriched gene sets between a pair of drugs, respectively. Performances of these scoring systems were evaluated by the probabilistic c-index (PC-index) devised by the DREAM consortium. We also applied the proposed method to the Connectivity Map dataset to explore more potential synergistic drug combinations.

Results: Using a gold standard derived by the DREAM consortium, we confirmed the prediction power of the three scoring systems (all \( P \)-values < 0.05). The co-gene/GS score achieved the best prediction of drug synergy (PC-index = 0.663, \( P \)-value < 0.0001), outperforming all methods proposed during DREAM challenge. Furthermore, a binary classification test showed that co-gene/GS scoring was highly accurate and specific. Since our method is constructed on a gene set-based analysis, in addition to synergy prediction, it provides insights into the functional relevance of drug combinations and the underlying mechanisms by which drugs achieve synergy.

Conclusions: Here we proposed a novel and simple method to predict and investigate drug synergy, and validated its efficacy to accurately predict synergistic drug combinations and to comprehensively explore their underlying mechanisms. The method is widely applicable to expression profiles of other drug treatments and is expected to accelerate the realization of precision cancer treatment.

Keywords: Synergy prediction, Drug combination, Gene set enrichment analysis

© 2016 The Author(s). Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.
Background

Development of effective treatments for cancers is an essential issue in clinical therapeutics. Thanks to the accumulated knowledge of cancer pathology, cancer treatments have been gradually shifting from the one-size-fits-all cytotoxic approach to the precision medicine that targets specific pathological features on a personalized basis [1]. Based on the reductionist “one gene-one disease” premise [2], antitumor drugs are designed to inhibit the growth of cancer cells by targeting essential genes or pathways with high specificity and efficacy, reducing damages to the normal cells. However, even with such advance in cancer therapeutics, some patients still suffer from refractory responses due to the development of drug resistance. Because of the complexity and heterogeneity of cancers, single drug alone may not be effective enough to completely and continuously suppress the activity of critical oncogenes or pathways. A common feature of drug resistance is a continuous activation of drug targets or their downstream signaling pathways [3]. On the other hand, combinatorial drug therapy may effectively circumvent the acquisition of drug resistance and optimize the efficacy of anticancer drugs.

Response to combinatorial drug therapy is optimized when a drug combination achieves greater (synergistic) effects than independent effects [1]. The synergy of drugs can be assayed by testing the inhibition of tumor cell growth by individual drugs and their combinations in vitro, followed by a mathematical formulation by Loewe additivity or Bliss independence [1, 2]. However, given the large number of drugs that are approved by FDA or under clinical trials, it seems to be impractical to experimentally test the synergy of all possible drug combinations, motivating the development of efficient computational methods for systematic screening and prediction of synergistic combinations.

Previous studies have proposed a handful of computational approaches to analyze high-throughput molecular datasets for predicting the synergy of drug combinations [4–6]. One of the computational approaches is devised based on the gene expression profiles achieved from treatments of individual drugs ([7–9]; reviewed in [2]). With the accumulation of gene expression profiles of drug treatments [10–12], the performance of such approach to model the underlying mechanisms of drug treatment can be improved. In a recent community computational challenge, namely the DREAM challenge (http://dreamchallenges.org/and https://www.synapse.org/#!Synapse:syn2785778/wiki/70252), 31 computational methods were developed to predict the synergistic effects of a total of 91 drug pairs by using gene expression profiles and evaluated against an experimental gold standard [13]. The promising results achieved by these methods highlight that in silico synergy prediction is possible and may greatly reduce the costs to screen for synergistic drug combinations. However, the accuracy of predictions still remains to be optimized [13].

The proposed study is motivated by the need for an improved prediction algorithm to systematically screen for synergistic drug combinations based on whole-genome expression profiles. According to the observations of previous in vitro studies, two synergistic drugs may target common signaling pathways to reinforce their individual effects [14, 15]. We hypothesized that drugs with synergistic effects perturb similar biological functions, or similar genes in a biological function in cells. To test the hypothesis, we designed three prediction scores based on the similarities in gene and/or functional level changes induced by two individual drug treatments. Here the functional changes were modeled by a gene set enrichment analysis, which summarizes expressional profiles to the functional level and uncovers systematic, even when subtle, changes in biological functions [16]. Specifically, we identified commonly enriched gene sets, measured by gene-set enrichment scores, between two drug treatments. Degree of overlapping of these gene sets and of the disturbed gene components in these gene sets between two drugs were used to construct three scoring systems. Performance of the scores was tested against the gold standard of 91 drug pairs provided by the DREAM challenge. In order to identify potential synergistic combinations over a broader range of drugs, we applied our method to the Connectivity Map (CMap) dataset [10], which includes the expression profiles associated with more than a thousand compounds, and identified both previously confirmed and novel synergistic drug pairs for breast cancer. Furthermore, since the gene sets represent biological functions, our prediction models also provide biological insights into the underlying mechanisms regulated by synergistic drugs.

Methods

Datasets from the DREAM challenge

We downloaded the expression profiles of Diffuse Large B-Cell Lymphoma OCI-LY3 cell line treated independently with 14 drugs at 2 different concentrations or DMSO (the control) at 6, 12, or 24 h after treatment from the Gene Expression Omnibus (accession ID GSE51068). The platform used to profile the gene expression was Affymetrix Human Genome U219 Array. We used the expression profiles obtained at 24 h because the IC_{20} values were measured only at 24 and 48 h in the original study [13].

A gold standard ranking of the 91 drug combinations was provided by the DREAM consortium to evaluate the performance of prediction methods. The drug combinations were ranked according to the value of excess over Bliss (EOB), the difference between the expected fractional inhibition and the induced fractional inhibition by
drug combination [13]. A pair of drugs with EOB \( \sim 0 \) have additive effect while positive (or negative) EOB values represent synergistic (or antagonistic) effects.

**CMap dataset**

A total of 2,588 expression profiles of a breast cancer cell line, MCF7, treated by 1,118 small-molecule perturbagens were downloaded from the CMap database [10].

**Gene sets**

Gene sets were downloaded from the BioMart website of Ensembl (http://www.ensembl.org/index.html) and the Molecular Signature database (MSigDB v3.1, http://www.broadinstitute.org/gsea/msigdb/index.jsp) [16]. Totally 5,357 gene sets of Gene Ontology biological process, molecular function, and cellular component, chemical and genetic perturbations, and oncogenic signatures were collected. To eliminate gene sets with highly overlapped gene components, we clustered these gene sets by kappa statistics and selected one gene set from each cluster as a representative gene set. A total of 2,181 representative gene sets were selected for subsequent analysis.

**Gene set enrichment analysis**

We started the analysis by performing gene set enrichment analysis for each drug in order to investigate the functions regulated by the drug. The enrichment score was calculated as

\[
S_l = \frac{1}{N_l} \sum_{j=1}^{N_l} Z_{ij}
\]

where \( N_l \) is the number of genes in gene set \( l \), and \( Z_{ij} \) is the delta \( z \)-score of gene \( j \) in sample \( i \). The delta-\( z \)-score was achieved by \( z \)-transforming the expression value of gene \( j \) in all samples \( (X_{ij}) \) followed by a subtraction of the \( z \)-transformed DMSO-treated normal control, as the equation below

\[
Z_{ij} = \frac{X_{ij} - \mu_j}{\sigma_j} - \frac{X_{DMSO-j} - \mu_j}{\sigma_j}
\]

where \( \mu_j \) stands for the mean of the gene \( j \) and \( \sigma_j \) is the standard deviation of the gene \( j \) among all samples. In order to evaluate the statistical significance of an enrichment score, we permuted the dataset for 1,000 times and calculated a \( P \)-value from the empirical distribution. Benjamini-Hochberg adjustment was performed to avoid multiple testing problems [17]. Adjusted \( P \)-value \( < 0.05 \) was set as a threshold to define significantly enriched gene sets. Non-informative gene sets, defined as those falling within the \( L_{0.05} \) range

\[
L_{0.05} = \pm 1.96 \frac{1}{\sqrt{N}}
\]

[18], where \( N \) is the number of genes in the gene set, were eliminated from subsequent analysis.

**Prediction scores**

Our hypothesis is that the synergy of two drugs (say, \( d_1 \) and \( d_2 \)) can be achieved by modulating common genes, similar biological functions, or common gene components within these biological functions. Accordingly, we devised three scoring methods. Suppose there are \( n \) commonly enriched gene sets between the two drugs. The first score (co-gene score) measures the co-disturbed genes between two drug treatments. The co-gene score for a pair of drugs \( d_1 \) and \( d_2 \) is

\[
\text{co-gene score}_{d_1,d_2} = \frac{g_{d_1,d_2}}{G}
\]

where \( G \) and \( g_{d_1,d_2} \) are the total number of genes and the number of commonly regulated genes with significant expression changes (with \( P \)-value for delta \( z \)-scores < 0.05) between \( d_1 \) and \( d_2 \). The second score (co-GS score) is constructed based on the similarity in enriched gene sets between the two drugs. Mathematical definition of the co-GS score is given by

\[
\text{co-GS score}_{d_1,d_2} = \frac{n_{d_1,d_2}}{L}
\]

with \( L \) representing the total number of gene sets, and \( n_{d_1,d_2} \) denoting the number of gene sets with significant enrichment in both drugs. The third score, namely the co-gene/GS score, is an average percentage of commonly disturbed genes within the co-enriched gene sets:

\[
\text{co-gene/GS score}_{d_1,d_2} = \frac{\sum_{l=1}^{N_{d_1,d_2}} n_l}{n_{d_1,d_2}}
\]

where \( N_{d_1,d_2} \) denotes the number of genes with significantly changes (with \( P \)-value for delta \( z \)-scores < 0.05) in both drug treatments within a co-enriched gene set \( l \) for \( d_1 \) and \( d_2 \).

**Performance evaluation**

To evaluate the performance of the proposed prediction scores, we employed the probabilistic c-index (PC-index) developed by the DREAM consortium [13], which was modified from the concordance index (c-index). The c-index computes the proportion of concordance between predicted and experimentally validated ranks by defining
a score \( t_{q,r} \) for quantifying the concordance for two pairs of drugs \( q \) and \( r \) \((q, r \in [1, 91])\). The score is given by

\[
t_{q,r} = \begin{cases} 
0, & \text{if } (u_q > u_r \& \& v_q < v_r \text{ or } u_q < u_r \& \& v_q > v_r) \\
1, & \text{if } (u_q > u_r \& \& v_q > v_r \text{ or } u_q < u_r \& \& v_q < v_r)
\end{cases}
\]

and the c-index is the average of all \( t_{q,r} \) given \( q \neq r \)

\[
c-index = \frac{\sum_{q=1 \ldots 90} \sum_{r=q+1 \ldots 91} t_{q,r}}{91 \choose 2}
\]

where \( u \) stands for the experimentally validated ranks and \( v \) is the predicted ranks. Considering the noise introduced during the experiment processes and high-throughput profiling, the DREAM consortium modified the scoring function by introducing an error function (erf) and computed the PC-index as follows

\[
t'_{q,r} = \begin{cases} 
\frac{1}{2} + \frac{1}{2} erf\left(\frac{EOB_q - EOB_r}{\text{sem}EBO_q^2 + \text{sem}EBO_r^2}\right), & \text{if } v_q < v_r \\
\frac{1}{2} - \frac{1}{2} erf\left(\frac{EOB_q - EOB_r}{\text{sem}EBO_q^2 + \text{sem}EBO_r^2}\right), & \text{if } v_q > v_r
\end{cases}
\]

and

\[
\text{PC-index} = \frac{\sum_{q=1 \ldots 90} \sum_{r=q+1 \ldots 91} t'_{q,r}}{91 \choose 2}
\]

where \( EOB_q \) is the excess over Bliss of a drug pair \( q \), provided in the gold standard profile, and \( \text{sem}EBO_q \) is the standard error of the mean of \( EOB_q \). When there is a concordance between predicted and experimentally validated ranks, the score \( t'_{q,r} \) falls between 0.5 and 1; otherwise, \( t'_{q,r} \) will be within \([0,0.5]\). The DREAM consortium tested the range of the PC-index in the gold standard dataset and found that the maximum PC-index \((\text{PC}_{\text{max}})\) was 0.90 and the minimum PC-index \((\text{PC}_{\text{min}})\) was 0.10; the normalized PC-index is defined as

\[
\text{PC-index}_{\text{norm}} = \frac{\text{PC-index} - \text{PC}_{\text{min}}}{\text{PC}_{\text{max}} - \text{PC}_{\text{min}}}.
\]

For the scoring method with the highest PC-index, we computed the area under ROC curve (AUC) and performed the precision analyses as described by Bansal et al. [13]. To calculate AUC, first, drug pairs with synergy and antagonism were defined from the gold standard profile. These drug pairs were identified by the DREAM consortium using the signal to noise ratio (SNR) computed as

\[
\text{SNR} = \frac{\text{mean} \ EOB}{\text{sem} \ of \ EOB}
\]

where sem is the standard error of the mean of the EOB. If the SNR of a drug pair is greater than 2 and its EOB is positive, it is defined as a synergistic drug pair. Oppositely, if the SNR of a drug pair is greater than 2 and its EOB is negative, it is defined as an antagonistic drug pair. Other drug pairs were defined as additive. In the original paper [13], sixteen drug pairs in the gold standard were defined as synergistic, 36 drug pairs as antagonistic, and 39 drug pairs as additive. We then ranked all drug pairs according to the co-gene/GS score in a decreasing manner. The true positive rate (TPR) and false positive rate (FPR) for synergy prediction were calculated as

\[
\text{TPR}_i = \frac{TP_i}{TP_i + FN_i}; \\
\text{FPR}_i = \frac{FP_i}{FP_i + TN_i};
\]

where \( TP_i, FN, FP_i, \) and \( TN_i \) denote the numbers of true positives, false negatives, false positives, and true negatives, respectively, given the top \( i \) drug pairs were called as synergistic. The \( TPR_i \) and \( FPR_i \) for antagonism predictions were calculated similarly, while the drug pairs were ranked based on the prediction scores in an increasing order. The results were visualized using ROC curves and corresponding AUC values were computed. Furthermore, we calculated the precision for synergy and antagonism prediction as

\[
\text{Precision(synergy)} = \frac{TP_{16}}{16}; \\
\text{Precision(antagonism)} = \frac{TP'_{36}}{36};
\]

where \( TP'_{36} \) is the number of true positives when the top 36 predicted drug pairs in the increasing order were called as antagonistic.

**Results**

**Model overview**

In the present study, we aim to test the hypothesis that similarities in expression functional changes induced by two drug treatments can predict the synergy of drugs. Figure 1 shows the overall design of this study. Based on the hypothesis, we devised three prediction scores. We tested performance of these scores by an experimentally assessed gold standard provided by the DREAM consortium, which was composed of a synergy ranking of 91 pairwise combinations among 14 well-known drugs. The rank of synergistic effects of the 91 combinations predicted by each score was compared to the gold standard
and the concordance of the ranks was measured by the PC-index (detailed in the Methods). In order to extend the scale of our prediction, we applied the best performing scoring method to the CMap dataset that comprises expression profiles of more than 1,000 compounds.

Figure 2 is the flowchart of the proposed scoring methods. The three scoring methods were constructed based on the similarities of co-enriched genes/gene sets (representing biological functions and pathways) and/or gene components within the gene sets. Specifically, the co-GS and co-gene methods measure the degree of overlap between enriched gene sets and disturbed genes, respectively, between two drugs. On the other hand, the co-gene/GS score focuses on the similarity (intersection) of genes disturbed by drug treatments within the co-enriched gene sets. We ranked all drug combinations
based on each prediction score; combinations with higher prediction scores are interpreted to have stronger synergistic effects, while low scores predict antagonistic effects. Mathematical definitions of the scores are provided in the Methods section. Since the gene sets represent biological functions, we were able to investigate the underlying mechanisms that account for the efficacy of drugs and the synergy of pairs of drugs.

Performance evaluation of the proposed scores by DREAM gold standard
We assessed the performance of the proposed scores against the DREAM gold standard. Notably, all of the three scores achieved significant PC-index (empirical \( P \)-value < 0.01; Table 1), confirming our hypothesis that synergistic drugs perturb similar genes and biological functions, as well as similar genes related to a biological function. The intersection between genes disturbed by each drug (co-gene score) achieved a PC-index of 0.648 (\( P \)-value < 0.0001), outperforming all the 31 community-generated approaches in DREAM challenge (range, 0.613 to 0.420; mean, 0.510). At the gene set level, degree of overlapping in gene sets enriched in two drug treatments (co-GS) was predictive of synergistic effect (PC-index = 0.589 and \( P \)-value = 0.0036). Furthermore, integrating the gene set and gene levels, the co-gene/GS score achieved even higher performance (PC-index = 0.663 and \( P \)-value < 0.0001). Taken together, we inferred that reinforced regulation of similar biological functions is a crucial mechanism for drug synergy. The observation is consistent with previous findings that synergistic drugs affect the expression of genes involved in common pathways [15, 19, 20].

Further investigation into the co-gene/GS score
The co-gene/GS score was the best-performing method among our three scores and outperformed all prediction models proposed during the DREAM challenge. We further investigated its prediction power. The score was significantly positively correlated with the excess over Bliss (EOB) values, an experimentally derived indicator of synergy (\( r = 0.57 \) and \( P \)-value = 5.01 \( \times \) 10^{-5}; Fig. 3a). Table 2 tabulates the best 15 predicted synergistic drug pairs. These top pairs showed significantly higher EOB than the others (\( t \)-test one-tailed \( P \)-value = 1.36 \( \times \) 10^{-4}; Fig. 3b). Among them, the top synergistic drug pairs were camptothecin & mitomycin C (score = 0.084), camptothecin & doxorubicin (0.036), and H-7 & mitomycin C (0.032). Taken together, the proposed co-gene/GS score was strongly correlated with the EOB values and accurately predictive of synergistic effects of drug pairs.

We further analyzed the performance of the co-gene/GS score by AUC and precision analyses. The ROC curves for synergy and antagonism prediction are shown in Fig. 4a and b, respectively. The AUC for synergy prediction was 0.87, which outperformed all the methods proposed during the DREAM challenge (range, 0.813 to 0.207; mean, 0.515) and was even higher than the integrated method proposed in [13]. However, the AUC for antagonism prediction (AUC = 0.36) was worse than random guess; AUC of the 31 previous methods ranged from 0.677 to 0.337 with an average of 0.517. The precision for synergy and antagonism predictions was 62.5 and 47.2 %, respectively, higher than random guess (17.6 and 40.0 %) [13]. Notably, the high precision of our method in predicting drug synergy again outperformed all the methods proposed in the DREAM challenge.

Since the score was designed based on a gene set analysis scheme, the method also reports the commonly enriched biological functions for each drug combination. Take the top predicted pair as an example, camptothecin and mitomycin C were commonly enriched in gene sets of HOOI_ST7_TARGETS_UP, AMUNDSON_DNA_DAMAGE_RESPONSE_TP53, NIKOLSKY_BREAST_CANCER_19Q13.1_AMPLICON, KUMAMOTO_RESPONSE_TO_NUTLIN_3A_UP, and ZHU_SKIL_TARGETS_UP. Detailed lists of co-enriched gene sets for all drug combinations are provided in Additional file 1: Table S1.

Application to the CMap dataset
After validating our method by the DREAM gold standard, we applied it to predict the synergy over a broader range of drugs analyzed in the CMap dataset. We used the expression profiles of MCF7 cell line treated with 1,118 compounds. A total of 2,588 instances (different compound-concentration combinations) were analyzed. Combinations of each drug pair at different concentrations were represented by the most synergistic one. The results of 1,038,710 drug combinations are shown in Additional file 2: Table S2. Top 10 synergistic pairs are listed in Table 3. These drug combinations are warranted candidates for further in vitro investigations.

Discussion
In the proposed study we designed three prediction models to fully test the hypothesis that synergy of drugs can be achieved by targeting common biological pathways.
or genes. The hypothesis was confirmed by using the DREAM in vitro dataset. Specifically, the co-gene/GS scoring system, measuring commonly gene-level disturbances within co-enriched gene sets between two drugs achieved high prediction accuracy, AUC, and precision for drug synergy and outperformed all proposed methods in the DREAM challenge.

Another advantage of our method is the ability to identify the underlying functions/mechanisms that lead to the efficacy of drug treatments and the synergy of drug combinations. For example, for the top synergistic drug pair reported in the DREAM gold standard, doxorubicin & H-7, we identified 14 commonly enriched gene sets (Table 4). One of the gene sets, “NEWMAN_ERCC6_TARGETS_DN”, consists of genes involved in DNA repair- and transcription-related pathways [21]. Both doxorubicin and H-7 are known to affect transcription in cells. Doxorubicin is widely used as a first-choice anticancer drug in a variety of tumors, including breast, lung, ovary, thyroid, and leukemia. It intercalates with DNA base pairs, binds to topoisomerase II, and causes DNA damages that may activate apoptotic pathways when the attempt to repair DNA breaks fails [22, 23]. In addition, doxorubicin has also been reported to directly affect DNA transcription via inducing histone eviction in promoter regions [24]. H-7 is well-characterized for its ability to de-reouple DNA transcription. It acts as a protein kinase C (PKC) inhibitor [25]. PKC family has been reported to affect the activity of some transcription factors and regulate gene transcription [26, 27]. As an inhibitor of PKC, treatment of H-7 affects the downstream pathways of PKC and thereby alters the transcription processes of cell.

Table 2 Top 15 synergistic drug pairs predicted by the co-gene/GS score in the DREAM dataset

| Drug pair                  | Co-gene/GS score | Predicted rank | Gold-standard rank |
|----------------------------|------------------|----------------|--------------------|
| Camptothecin & Mitomycin C | 0.084            | 1              | 5                  |
| Camptothecin & Doxorubicin | 0.036            | 2              | 16                 |
| H-7 & Mitomycin C          | 0.032            | 3              | 2                  |
| Methotrexate & Mitomycin C | 0.030            | 4              | 28                 |
| Doxorubicin & Mitomycin C  | 0.027            | 5              | 4                  |
| Cycloheximide & H-7        | 0.027            | 6              | 9                  |
| Camptothecin & Etoposide   | 0.025            | 7              | 15                 |
| H-7 & Trichostatin A       | 0.019            | 8              | 19                 |
| H-7 & Rapamycin            | 0.017            | 9              | 27                 |
| Camptothecin & H-7         | 0.015            | 10             | 10                 |
| Etoposide & Mitomycin C    | 0.015            | 11             | 3                  |
| H-7 & Vincristine          | 0.012            | 12             | 43                 |
| H-7 & Monastrol            | 0.012            | 13             | 14                 |
| Cycloheximide & Methotrexate| 0.011           | 14             | 64                 |
| Doxorubicin & H-7          | 0.011            | 15             | 1                  |
Taken together, both these two drugs can change transcription activity, which is consistent with our findings that the treatments of these two drugs co-regulate the gene set, "NEWMAN_ERCC6_TARGETS_DN", composed of transcription-related genes.

Applying the prediction model to a huge collection of drug-treatment profiles deposited in CMap, we successfully identified potential synergistic drug combinations (Table 3). Taking the ninth pair, genistein & etoposide, as an example, the two drugs showed a similar enrichment in the gene set “peptide transporter activity” (Table 5), indicating their co-regulation of transporters such as MRPI (ABCC1), an drug-efflux pump causing cancer cell to attain resistance to some antitumor drugs [28]. Genistein, a natural isoflavonoid compound appearing in citrus fruits and soybean, has been reported to have antitumor effect on several types of cancers, including breast cancer [29]. It is also capable of inhibiting the multidrug resistance resulting from the well-known drug efflux proteins, such as MRPI, P-glycoprotein, and other ABC (ATP-binding cassette) proteins [30–33]. Etoposide, on the other hand, is a topoisomerase II inhibitor and has been reported to induce apoptosis in MCF7 cell line [34]. Being a substrate of MRPI, P-glycoprotein and other multidrug resistance-associated proteins, etoposide is ineffective to suppress tumor growth when these proteins are largely produced [35, 36]. Previous in vitro studies showed that genetic deletion of the MRPI proteins greatly restores the sensitivity of tumor cells to etoposide due to the minor contributions of other drug-efflux proteins, such as P-glycoprotein, to the drug resistance of etoposide [36, 37]. Taken together, 

**Fig. 4** Receiver operating characteristic (ROC) curves for drug synergy and antagonism prediction. a ROC curve for drug synergy prediction. The area under the ROC curve (AUC) is 0.87. b ROC curve for drug antagonism prediction. The AUC is 0.36

**Table 3** Top 10 predicted drug combinations in the CMap dataset

| Drug combination | Co-gene/GS score | Rank |
|------------------|------------------|------|
| Prestwick-682 & MG-262 | 0.667 | 1 |
| Cefazolin & Nocodazole | 0.667 | 2 |
| Anisomycin & Prednisolone | 0.625 | 3 |
| Lisinopril & Suramin sodium | 0.600 | 4 |
| Iopanoic acid & Butacaine | 0.563 | 5 |
| Alpha-ergocryptine & Clofazimine | 0.529 | 6 |
| Alpha-ergocryptine & Diloxanide | 0.524 | 7 |
| LM-1685 & Mepyramine | 0.512 | 8 |
| Genistein & Etoposide | 0.504 | 9 |
| Acetohepxamide & Benztiazide | 0.500 | 10 |

**Table 4** Commonly enriched gene sets of doxorubicin and H-7 in the DREAM dataset

| Gene set | MSigDB category |
|----------|-----------------|
| NEWMAN_ERCC6_TARGETS_DN | Chemical and genetic perturbations |
| LEE_NEURAL_CREST_STEM_CELL_DN | Chemical and genetic perturbations |
| ODNNEIN_METASTASIS_UP | Chemical and genetic perturbations |
| CERVERA_SDHB_TARGETS_2 | Chemical and genetic perturbations |
| OSADA_ASCL1_TARGETS_UP | Chemical and genetic perturbations |
| YAUCH_HEDGEHOG_SIGNALING_PARACRINE_UP | Chemical and genetic perturbations |
| HINATA_NFkB_TARGETS_KERATINOCYTE_DN | Chemical and genetic perturbations |
| TAVAZOE_METASTASIS | Chemical and genetic perturbations |
| VALK_AML_CLUSTER_7 | Chemical and genetic perturbations |
| MIKELSEN_NPC_ICP_WITH_H3K27ME3 | Chemical and genetic perturbations |
| HOELZEL_NF1_TARGETS_UP | Chemical and genetic perturbations |
| PTEN_DN.V2_UP | Oncogenic signatures |
| KRAS.DF.V1_DN | Oncogenic signatures |
| Response to wounding | Gene ontology terms |
when these two drugs are combined, the inhibitory effect on multidrug resistance of genistein may sensitize cellular responses to etoposide, thus intensifying the antitumor effects through apoptosis.

There are some limitations in our methods. First, the gene set enrichment score was calculated by summing all the expression values, thus the topological relationships between genes were not considered. This may lead to an underestimation of those upstream genes that are able to cause significant disturbance pathway-wise even when they are slightly changed. Future studies may extend this work by using other gene set scoring methods that weight content genes based on topological information [38, 39] and significance of changes [16, 40]. Besides, this scoring scheme ignores the directional (i.e., upward and downward) changes in genes of a gene set. This may cause biases in estimating the influence of expression changes on biological functions. Thirdly, in the calculation of prediction scores we considered a simple intersection, but ignored the directional agreement of changes in two gene sets or two genes. Further improvement to our method that tackles this concern may provide better biological insights. Fourthly, our proposed similarity-based scoring methods were highly predictive of drug synergy, but did not performed well in predicting drug antagonism. This was consistent to the findings from the DREAM challenge: hypotheses needed to predict synergy and antagonism may be quite different [13]. Furthermore, drug pairs could attain synergistic/antagonistic effects by varied mechanisms. Our proposed method, even though proved to have high prediction power, may only address one specific type of mechanisms. To fully explore all possibilities and further improve the robustness of the prediction method, future studies may integrate diverse data types of the treated cells, such as methylation and gene mutation data, and combine different prediction models, as well as other hypothesis.

Conclusions
In this paper, we comprehensively tested and confirmed the hypothesis that synergy of drugs can be achieved by regulating common biological functions and genes. A synergy-predicting score was proposed and validated by an experimentally assessed gold standard. The prediction performance of this simple score was better than previous methods. We applied our devised method to a larger collection to screen for potential drug combinations. We also demonstrated that the method not only achieves accurate prediction but also investigates the underlying mechanisms of drug synergy.

Additional files

Additional file 1: Table S1. List of co-enriched gene sets for all drug combinations in the DREAM dataset. (XLSX 35 kb)
Additional file 2: Table S2. Prediction of all 1,038,710 drug combinations in the CMap dataset. (XLSX 28927 kb)

Acknowledgements
The study was supported by the National Health Research Institutes of Taiwan (NHRI-EX104-10419B and NHRI-EX105-10419B), the YongLin Healthcare Foundation (FB002-7), and the Ministry of Science and Technology of Taiwan (103–2917-I-002–166). The authors thank the Center of Genomic Medicine, National Taiwan University, for providing financial support and computing facilities. The authors also greatly appreciate the brilliant and constructive inputs from the reviewers and participants of the International Conference on Intelligent Biology and Medicine (ICIBM 2015).

Declarations
Publication charges for this article have been funded by the National Health Research Institutes of Taiwan (NHRI-EX104-10419B and NHRI-EX105-10419B) and the YongLin Healthcare Foundation (FB002-7).
This article has been published as part of BMC Systems Biology Volume 10 Supplement 3, 2016: Selected articles from the International Conference on Intelligent Biology and Medicine (ICIBM 2015). systems biology. The full contents of the supplement are available online at http://bmcsystbiol.biomedcentral.com/articles/supplements/volume-10-supplement-3.

Availability of data and materials
All microarray data used in this study are publicly available: DREAM dataset, http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE51068; Cmap dataset, https://www.broadinstitute.org/cmap/.

Authors’ contributions
All authors conceived the study together. YH, YuC, and TH designed the analysis pipeline. YH and YuC carried out the data analysis and drafted the manuscript. YIC, TH, and EYC revised the manuscript. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Consent for publication
Not applicable.

Ethics approval and consent to participate
Not applicable.

Author details
1 Graduate Institute of Biomedical Electronics and Bioinformatics, National Taiwan University, Taipei, Taiwan. 2Greehey Children’s Cancer Research Institute, University of Texas Health Science Center at San Antonio, San Antonio, Texas, USA. 3Department of Epidemiology and Biostatistics, University of Texas Health Science Center at San Antonio, San Antonio, Texas, USA. 4Department of Medical Research, Taichung Veterans General Hospital, Taichung, Taiwan. 5Bioinformatics and Biostatistics Core, Center of Genomic Medicine, National Taiwan University, Taipei, Taiwan.

Published: 26 August 2016

References
1. Al-Lazikani B, Banerji U, Workman P. Combinatorial drug therapy for cancer in the post-genomic era. Nat Biotechnol. 2012;30(7):679–92.
2. Ryall KA, Tan AC. Systems biology approaches for advancing the discovery of effective drug combinations. Journal of cheminformatics. 2015;7:7.

3. Garraway LA, Janne PA. Circumventing cancer drug resistance in the era of personalized medicine. Cancer discovery. 2012;2(3):24–26.

4. Xu KJ, Song J, Zhao XM. The drug cocktail network. BMC Syst Biol. 2012;6 Suppl 1:S5.

5. Huang L, Li F, Sheng J, Xia X, Ma J, Zhan M, Wong ST. DrugComboRanker: drug combination discovery based on target network analysis. Bioinformatics. 2014;30(12):i228–236.

6. Tang J, Kirihimel L, Xu T, Sawajda A, Yadav B, Wenneberger K, Alttokallo T. Target inhibition networks: predicting selective combinations of druggable targets to block cancer survival pathways. PLoS Comput Biol. 2013;9(9):e1003226.

7. Yang J, Tang H, Li Y, Zhong R, Wang T, Wong S, Xiao G, Xie Y. DIGRE: Drug-Induced Genomic Residual Effect model for successful prediction of multidrug effects. CPT: Pharmacometrics & Systems Pharmacology. 2015;4:e1.

8. Goswami CP, Cheng L, Alexander P, Singal A, Li L. A New drug combinatory effect prediction algorithm on the cancer cell based on gene expression and dose-response curve. CPT: Pharmacometrics & Systems Pharmacology. 2015;4:e9.

9. Zhao J, Zhang XS, Zhang S. Predicting cooperative drug effects through the quantitative cellular profiling of response to individual drugs. CPT Pharmacometrics Syst Pharmacol. 2014;3:e102.

10. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, Lerner J, Brunet JP, Subramanian A, Ross KN et al. The connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science. 2006;313(5795):1929–35.

11. Kim J, Yoo M, Kang J, Tan AC. K-Map: connecting kinases with therapeutics. Proteomics. 2013;13(1):214–21.

12. Yoo M, Shin J, Kim J, Ryall KA, Lee K, Lee S, Jeon M, Kang J, Tan AC. DsigDB: drug signature databases for gene set analysis. Bioinformatics. 2015.

13. Pietro AD, Conseil G, Perez-Victorian, JM, Dayan G, Baubichon-Cortay H, Trompier D, Steinfields E, Jault JM, Wet HD, Maitejne M et al. Modulation by flavonoids of cell multidrug resistance mediated by P-glycoprotein and related ABC transporters. Cell Mol Life Sci. 2002;59:16.

14. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and unified approach. Journal of education. 1995;57(1):289–301.

15. Newman JC, Bailey AD, Weiner AM. Cockayne syndrome group B protein (CSB) plays a general role in chromatin maintenance and remodeling. Proc Natl Acad Sci U S A. 2006;103(23):8963–8.

16. Tarca AL, Draghici S, Khatri P, Amin N, Amini A, Voicu C, Georgescu C, Romero R. A systems biology approach for pathway level analysis. Genome Res. 2007;17(10):1537–45.

17. Benjamin Y-H, Hidaka H. Pharmacology of protein kinase inhibitors. Annu Rev Pharmacol Toxicol. 1992;32:21.

18. Lingman FA, Hidaka H, Nishida E, Noda M, Hidaka H. Mitochondria and the drug target: the role of the outer membrane. Biochim Biophys Acta. 1990;1036(1):1–5.

19. Inaba M, Hidaka H, Nishida E. Pharmacology of protein kinase inhibitors. Annu Rev Pharmacol Toxicol. 1994;34:167–92.

20. Pan C, Zhai Y, Wang L, Zhang Y, Liu J, Wang J, Li W, Wang Z, Cao X, Jin X et al. A novel significance score for gene selection and ranking. Bioinformatics. 2014;30(6):781–7.

21. Litman TA, Druey JB, Stein WD, Bates SE. From MDR to MXR: new strategies for drug resistance. Curr Opin Pharmacol. 2002;2(3):214–21.

22. Tan AC, Bailey AD, Weiner AM. The cocktail network. BMC Syst Biol. 2012;6 Suppl 1:S5.

23. Hidaka H, Kobayashi R, Pharmacology of protein kinase inhibitors. Annu Rev Pharmacol Toxicol. 1992;32:21.

24. Cai GC, von Burstina VA, Lopez-Haber C, Kazanietz MG. Differential regulation of gene expression by protein kinase C isoforms as determined by genome-wide expression analysis. J Biol Chem. 2011;286(13):11254–64.

25. Ray A, Fields AP, Ray BK. Activation of transcription factor SAF involves its phosphorylation by protein kinase C. J Biol Chem. 2000;275(50):39727–33.

26. Effert T. The human ATP-binding cassette transporter genes: from the bench to the bedside. Current Molecular Medicine. 2001;1:11.

27. Peterson G, Bames S. Genistein inhibition of the growth of human breast cancer cells: independent from estrogen receptors and the multidrug resistance gene. Biochim Biophys Res Commun. 1991;179:7.

28. Trompier D, Baubichon-Cortay H, Chang XB, Maitejne M, Baron D, Riordon JR, Di Pietro A. Multiple flavonoid-binding sites within multidrug resistance protein MRP1. Cellular and molecular life sciences : CMLS. 2003;60(10):214–77.

29. Lane H. The cell cycle and we will help you at every step: Submit your next manuscript to BioMed Central and we will help you at every step:

30. Submit your next manuscript to BioMed Central and we will help you at every step:

31. We accept pre-submission inquiries

32. Our selector tool helps you to find the most relevant journal

33. We provide round the clock customer support

34. Convenient online submission

35. Thorough peer review

36. Inclusion in PubMed and all major indexing services

37. Maximum visibility for your research

38. Submit your manuscript at www.biomedcentral.com/submit