DrugComboRanker: drug combination discovery based on target network analysis

Lei Huang\(^1,2,\dagger\), Fuhai Li\(^2,\dagger\), Jianting Sheng\(^2\), Xiaofeng Xia\(^2\), Jinwen Ma\(^1\), Ming Zhan\(^2\) and Stephen T.C. Wong\(^2\)*

\(^1\)Department of Information Science, School of Mathematical Sciences and LMAM, Peking University, Beijing, 100871, China and \(^2\)NCI Center for Modeling Cancer Development, Department of Systems Medicine and Bioengineering, Houston Methodist Hospital Research Institute; Weill Cornell Medical College of Cornell University, Houston, TX 77030, USA

Abstract

**Motivation:** Currently there are no curative anticancer drugs, and drug resistance is often acquired after drug treatment. One of the reasons is that cancers are complex diseases, regulated by multiple signaling pathways and cross talks among the pathways. It is expected that drug combinations can reduce drug resistance and improve patients’ outcomes. In clinical practice, the ideal and feasible drug combinations are combinations of existing Food and Drug Administration-approved drugs or bioactive compounds that are already used on patients or have entered clinical trials and passed safety tests. These drug combinations could directly be used on patients with less concern of toxic effects. However, there is so far no effective computational approach to search effective drug combinations from the enormous number of possibilities.

**Results:** In this study, we propose a novel systematic computational tool DrugComboRanker to prioritize synergistic drug combinations and uncover their mechanisms of action. We first build a drug functional network based on their genomic profiles, and partition the network into numerous drug network communities by using a Bayesian non-negative matrix factorization approach. As drugs within overlapping community share common mechanisms of action, we next uncover potential targets of drugs by applying a recommendation system on drug communities. We meanwhile build disease-specific signaling networks based on patients’ genomic profiles and interactome data. We then identify drug combinations by searching drugs whose targets are enriched in the complementary signaling modules of the disease signaling network. The novel method was evaluated on lung adenocarcinoma and endocrine receptor positive breast cancer, and compared with other drug combination approaches. These case studies discovered a set of effective drug combinations top ranked in our prediction list, and mapped the drug targets on the disease signaling network to highlight the mechanisms of action of the drug combinations.

**Availability and implementation:** The program is available on request.

**Contact:** stwong@tmhs.org

*To whom correspondence should be addressed.

\(^\dagger\)The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

1 INTRODUCTION

Cancers are complex diseases regulated by interactions of multiple signaling pathways interacting with each other. Though there are some anticancer drugs, they are often not curative and associated with acquired drug resistance. It is believed that a single drug can target only one disease signaling module, whereas alternative signaling pathways will be activated to maintain tumor development. For example, in melanoma, the selective BRAF inhibitor, PLX4032, can induce tumor responses in the majority of patients with BRAF V600E mutant melanoma, but the duration of this response is limited owing to the development of acquired resistance, which may be due to alternative pathways parallel to the ‘BRAF-MEK-ERK’ signaling pathway (Johannsen et al., 2013; Nazarian et al., 2010). To reduce the drug resistance effect, and improve cancer treatment outcomes, drug combinations are considered as an optimal option.

The ideal drug combinations expected by clinicians are combinations of Food and Drug Administration (FDA)-approved drugs or existing bioactive compounds that have entered clinical trials and passed safety tests. These drug combinations could be used by patients without toxic side effects. Drug combination prediction has been a challenging task in computational biology, despite a few studies in the area. The genetic algorithm was, for example, proposed to find the optimal combinations of a small number of drugs based on the experimental results (Zinner et al., 2009). The method, however, is not able to identify optimal drug combinations from a large set of drugs, as it requires experimental validations in the prediction process. The machine learning-based approach was proposed to classify drugs into combinations based on molecular and pharmacological features of drugs (Zhao et al., 2011). However, the features of drugs are too general to be informative, and the method does not integrate the rich genomic profiles of drugs and disease into consideration.

Combinatorial Drug Assembler (CDA) was developed to identify drug combinations based on enrichment analysis of genomic profile data. However, the enrichment analysis implemented in the method is based on differentially expressed genes, disease driver genes are not clearly defined, so that drugs targeting the causal dysfunctional signaling modules of disease maybe missed out (Lee et al., 2012). To overcome the aforementioned problems, we propose a novel computational tool, DrugComboRanker, for predicting drug combinations targeting multiple signaling modules of cancer-specific networks through integrative analysis of genomic profiles of both drugs and cancers.

© The Author 2014. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
2 MATERIALS AND METHODS

2.1 Overview of DrugComboRanker

This study seeks to answer a clinical question that which drugs can be combined with existing therapy to reduce drug resistance and improve patients’ outcomes. We have thus developed a novel tool, DrugComboRanker, to prioritize drug combinations and uncover the underlying mechanisms of action. The approach is designed based on the following two observations. First, dysfunctional signaling networks of diseases are complex and stable, and effective drug combinations can inhibit major modules of the networks simultaneously. Second, drugs often have multiple active target genes or proteins. Grouping targets of drugs with similar transcriptional responses profiles allows better recovery of their targeted signaling modules in disease than based on individual targets alone. Figure 1 is an overview of the proposed DrugComboRanker. The drug functional network is firstly reconstructed based on genomic profiling data of drugs that are available in the Connectivity Map (CMAP) database. The Bayesian non-negative matrix factorization with the β-divergence (BNMFβD; Tan and Fevotte, 2013) approach is conducted to partition the drug functional network into connected functional communities, whose targets indicate the inhibited signaling modules of diseases. On the other hand, the dysfunctional signaling modules of diseases are reconstructed by integrating disease genomics data and protein interactome data. Drug combinations are then ranked from different drug functional communities, which inhibit disease signaling network modules.

2.2 Drug functional network reconstruction

The aim of drug function network reconstruction is to identify drug communities that share common transcriptional responses to drug treatment. We reconstruct the drug functional network based on the data of the CMAP (CMap, build 02; Lamb et al., 2006), which consists of 6100 gene expression profiles of four cancer cell lines (MCF7, PC3, HL60 and SKMEL5) treated by 1308 drugs at different doses. We first combine the gene expression profiles of four cancer cell lines (MCF7, PC3, HL60 and the STITCH database (Kuhn et al., 2006), respectively. The size of these signatures is heuristically determined as described in (Iorio et al., 2010). Finally, the gene set enrichment analysis (Subramanian et al., 2005) score, $S(i,j)$, is used as the dissimilarity metric between drug $i$ and drug $j.$ A threshold $T = 0.7848$ (the third quantile of the empirical probability distribution of the drug dissimilarity metric) is given to remove the non-significant dissimilarities as follows:

$$
S(i,j)_G =
\begin{cases} 
S(i,j), & \text{if } S(i,j) \leq T \\
C, & \text{else} 
\end{cases}
$$

Then we convert the dissimilarity into the normalized similarity score, $S_G(i,j)$ as $S_G(i,j) = 1 - \frac{NORM(S(i,j))}{\max_i(S(i,.))}$ where $S_G(i,.)$ denotes all the connection scores linking to the $i$-th node. Here, we empirically set $C = 10.$ In addition to the similarity of transcriptional responses, a different similarity metric, $S(i,j)$, was defined in STITCH database (Kuhn et al., 2012) by integrating a combined score of structure similarity, experimental similarity and text mining similarity score. The structure similarity is defined by the Tanimoto 2D chemical similarity scores (Martin et al., 2002). The experimental similarity is calculated by the Pearson correlation of the activity patterns of the compounds based on NCI60 screens. The text mining similarity is computed by mining the curated database, such as OMIM and MEDLINE, using a co-occurrence scheme and a natural language processing approach (Jensen et al., 2006; Šarić et al., 2006). In this study, we combine the two similarity metrics as $S(i,j) = S_G(i,j) + S(i,j).$ The rational to use the combined similarity is to generate better drug community structure when applying the clustering algorithm on the drug functional network. Finally, a drug functional network is reconstructed with 1308 nodes and 51 164 edges. Figure 2 shows an example of the reconstructed drug functional network community.

2.3 Drug community discovery by using Bayesian non-negative matrix factorization with β-divergence

Next, we partition the reconstructed drug functional network into communities (modules) within which drugs share common mechanisms of action (targeting the same or related signaling modules), as illustrated in Figure 2. We use the BNMFβD (Tan and Fevotte, 2013) for the network partition. The algorithm has shown good partition results on various classes of real-world signals such as audio and images (Tan and Fevotte, 2013). Specifically, the algorithm can (i) enable soft partitioning (one drug can be partitioned into multiple communities with different membership scores); (ii) reduce noise influence (suppresses noise in the high dimensional space); and (iii) allow multi-scale partition. To the best of our knowledge, the algorithm is for the first time used in drug functional network partition.

Specifically, given a weighted undirected network with $N$ nodes, let $V \in \mathbb{R}^{N \times N}$ denote its adjacent matrix. BNMFβD factorizes the matrix $V$ into two low-rank matrices:

$$
V \approx \tilde{V} = WH
$$

where $W$ and $H$ are non-negative matrices with dimensions $N \times K$ and $K \times N,$ and the elements of $V,$ $W,$ and $H$ are denoted by $v_{ij}, w_k,$ and $h_k,$ respectively. The $k$-th column of $W$ is denoted by $w_k \in \mathbb{R}^N,$ and the $k$-th row of $H$ is denoted by $h_k \in \mathbb{R}^N.$ The $W$ is normalized to $W_N,$ that

![Fig. 1. Overview of the DrugComboRanker for prioritizing disease-specific drug combinations](image)
Maximizing the posterior density is equivalent to minimizing the negative log posterior, which can be regarded as a loss function $C_{MAP}(W, H, \lambda)$ as

$$C_{MAP}(W, H, \lambda) = -\log p(W, H, \lambda | V)$$

$$= - \log p(V | W, H) - \log p(W | \lambda) - \log p(H | \lambda) - \log p(\lambda)$$

where $\log p(V | W, H)$ is the log-likelihood.

The generalized $\beta$-divergence is defined by

$$D_\beta(p(x) | q(x)) = \int p(x) \log \left( \frac{p(x)}{q(x)} \right) dx$$

$$= \int x \log \left( \frac{x^\beta}{q(x)} \right) dx - \beta \int \log q(x) dx$$

$$= x^\beta \log \frac{x}{\beta} - x^\beta + \frac{x^\beta}{\beta} - \beta \log \beta$$

The $\beta$-divergence can be regarded as a minus log-likelihood for the Tweedie distribution and its probability density function is given by

$$f(x; \mu, \phi, \beta) = h(x, \phi) \exp \left( \frac{1}{\phi} \left( \frac{1}{\beta - 1} x^\beta - \frac{1}{\beta} \mu^\beta \right) \right)$$

where $h(x, \phi)$ is the base measure function, $\mu$ is the mean, $\phi$ is the dispersion parameter and $\beta$ is the shape parameter. Assuming that $v_{ij}$ is generated from the Tweedie distribution, the log-likelihood function can be given by

$$-\log p(V | W, H) = \frac{1}{\phi} D_\beta(V | WH) + C$$

To insure $W$ and $H$ are non-negative, the Half-Normal priors are assigned on them,

$$p(w_{ij} | \lambda_i) = \mathcal{HN}(w_{ij} | \lambda_i)$$

$$p(h_{ij} | \lambda_i) = \mathcal{HN}(h_{ij} | \lambda_i)$$

and place an inverse Gamma priors on each $\lambda_i$,

$$p(\lambda_i; a, b) = \frac{b^a}{\Gamma(a)} \lambda_i^{-(a+1)} \exp(-\frac{b}{\lambda_i})$$

Then, according to Equation (5), the objective function $C_{MAP}(W, H, \lambda)$ can be given as

$$C_{MAP}(W, H, \lambda) = \frac{1}{\phi} D_\beta(V | WH) + \sum_{i=1}^K \frac{1}{\lambda_i} \left( \frac{1}{2} w_{ij}^2 + \frac{1}{2} h_{ij}^2 + b \right)$$

$$+ (N + a + 1) \log \lambda_i + C$$

To minimize $C_{MAP}(W, H, \lambda)$ with respect to $W$, $H$ and $\lambda$, we adopt the strategy in (Tan and Fevotte, 2013) by introducing a local majorization–minimization algorithm with efficient multiplicative updates. Finally, we give the overlapping community detection algorithm as follows:

Step 1: Initialize $w_{ij} \in \mathbb{R}_{+}^{K \times N}$ and $h_{ij} \in \mathbb{R}_{+}^{K \times N}$ to random non-negative values.

Step 2: Update $W$, $H$ and $\lambda_i$ by

$$H = H \cdot \left( \frac{W^T((WH)^{(\beta-2)} \cdot V) \cdot (WH)^{(\beta-1)} + \phi \cdot \text{repmat}(\lambda, 1, N)}{W^T(WH)^{(\beta-1)} \cdot (WH)^{(\beta-1)} + \phi \cdot \text{repmat}(\lambda, 1, N)} \right)^{\gamma(\beta)}$$

$$W = W \cdot \left( \frac{((WH)^{(\beta-2)} \cdot V)H^T \cdot (WH)^{(\beta-1)} + \phi \cdot \text{repmat}(\lambda, 1, N)}{((WH)^{(\beta-2)} \cdot V)H^T + \phi \cdot \text{repmat}(\lambda, 1, N)} \right)^{\gamma(\beta)}$$

Fig. 2. Snapshots of the drug functional network (top) and an example of the reconstructed drug network community (bottom)
where the transition matrix from drug based algorithm is designed based on a bipartite network projection technique. Let $\gamma$ be the known drug targets. The drug–target interaction network can be described as a bipartite graph $G=(X,Y,e)$. The elements of $X$ denote the differentially expressed genes of the gene expression profile of a disease. The enriched modules are then considered as disease susceptibility major components.

2.4 Drug combination discovery based on target network analysis

The novel drug combination approach consists of the following three major components.

(a) Disease specific signaling network reconstruction

Several approaches (Barrenas et al., 2012; Chuang et al., 2007; Ideker et al., 2002) have been proposed to reconstruct signaling networks of diseases based on transcriptome and interactome data. In this study, the approach proposed in (Barrenas et al., 2012) is used. The integrated protein–protein interactions from BioGRID (Stark et al., 2006) and the manually curated human cancer signaling from (Awan et al., 2007; Cui et al., 2007; Li et al., 2012; Newman et al., 2013) (available at http://www.bri.mrc.ox.ac.uk/wang/) are clustered into functional protein–protein modules, and each module is tested for enrichment (Fisher’s exact test, $P<0.01$) of the differentially expressed genes of the gene expression profile of a disease. The enriched modules are then considered as disease susceptibility modules, from which the highly interconnected genes are identified as the disease-specific signaling network. Figure 3 shows the reconstructed signaling network of lung adenocarcinoma.

(b) Functional drug target prediction using network-based recommendation

When two drugs are considered, $\Gamma(i,j)=k(t_i)\frac{1}{k(t_j)}$, and $k(t)$ is the degree of the $t$ node in the bipartite network. Targets with recommendation scores greater than given threshold, 0.1, are kept as the active functional targets of given drugs.

(c) Disease-specific drug combination discovery

With the constructed disease signaling network and the predicted drug targets, drug combinations are then prioritized by combining the following synergistic scores. Given two candidate drugs, $d_i$ and $d_j$, from different clusters, suppose $d_i\in C_k$ and $d_j\in C_n$, and $T_k = \{t_{k1}, t_{k2}, \ldots, t_{km}\}$ denote the targets of $d_i$ in $C_k$, and $T_n = \{t_{n1}, t_{n2}, \ldots, t_{nm}\}$ denote the targets of $d_j$ in $C_n$. The first synergistic score is defined as follows.

$$
S_1(i,j) = \frac{\sum_{t_{ki}\in T_k} \sum_{t_{mj}\in T_n} CS(t_{ki}, t_{mj}) \exp(-D(t_{ki}, t_{mj})/n^2)}{\sum_{t_{ki}\in T_k} \sum_{t_{mj}\in T_n} CS(t_{ki}, t_{mj})}
$$

where $CS(t_{ki}, t_{mj})$ is the centrality score of target $t_{ki}$ in the reconstructed disease signaling network, and it is an additive betweenness ($B_h$), closeness ($C_h$) (Brandes and Fleischer, 2005) and PageRank ($P_h$) score (Page et al., 1999) of protein $t_{ki}$, that is

$$
CS(t_{ki}) = B_h(t_{ki}) + C_h(t_{ki}) + P_h(t_{ki})
$$

These are three different but correlated centrality measurements, and the reason of combing them is to get a robust centrality score. The minimum shortest path from $t_{ki}$ to $T_n$ is the minimum shortest path from $t_{ki}$ to $T_k$. The first synergistic score, $S_1(i,j)$, prefers drug combinations, whose targets are in the center (hubs) of disease signaling network and closely connected.

The second synergistic score is defined as

$$
S_2(i,j) = \sum_{t_{ki}\in T_k} \sum_{t_{mj}\in T_n} \frac{\text{Sim}(t_{ki}, t_{mj})}{(m+n)(m+n-1)}
$$
where $\text{Sim}(t_i, t_j)$ is the semantic similarity of gene ontology (GO) annotations of $t_i$ and $t_j$ (Couto et al., 2007), which is computed based on the overlap of the GO terms that are associated with $t_i$ and $t_j$, and is defined as

$$
\text{Sim}(t_i, t_j) = \frac{2 \log_{\text{max}} \left[ p(A) \right]}{\log_{p} p(GO_i) + \log_{p} p(GO_j)}
$$

(19)

Where $GO_i$ is the GO term that associated with $t_i$, and $A$ is a GO term that is an ancestor of both $GO_i$ and $GO_j$, and

$$
p(GO_j) = \frac{\text{Freq}(GO_j)}{\text{MaxFreq}}
$$

(20)

$\text{Freq}(GO_j)$ is the frequency of GO term $GO_j$ occurring in GO annotations, which are taken from GO database. $\text{MaxFreq}$ is the maximum occurrences frequency of GO terms that are associated with all the targets and the predicted drug targets in the GO annotations. The second synergistic score, $S_2(i, j)$, prefers drug combinations that block genes with similar functions, e.g. cell proliferation.

Finally, the synergistic score of drug $d_i$ and $d_j$ is given by

$$
S_d(i, j) = (S_1(i, j) + S_2(i, j)) \cdot d_g(i, j)
$$

(21)

where

$$
d_g(i, j) = \max \left[ S_3(i, j), \sqrt{S_3(i, j) \cdot S_4(i, j) + S_i(i, j)} \right]
$$

(22)

where $S_3(i, j)$ reflects the distance of their expression pattern. In summary, drug combinations targeting on the disease-specific signaling network, with similar functions, through alternative targets are prioritized.

3 RESULTS

We have applied the BNMFβD algorithm to cluster the drug–drug network into overlapping drug communities. Table 1 lists the parameters in the analysis. We set $a$ as one of $\{10, 50, 100, 150, 200, 250, 300, 350, 400, 450\}$, and $b$ to be equal to $a$. When $a$ and $b$ are set to be 450, the BNMFβD algorithm will converge more quickly. Under this setting of $a$ and $b$, we test the drug network community reconstruction results with different $K$ and $\beta$, as $K = \{100, 110, 120, 130, 140, 150, 160, 170\}$ and $\beta = \{1, 2\}$. When $K$ is 150 and $\beta$ is 1, we get better drug community structures in terms of the following defined community quality measurements. We assign the number from 1 to 150 to each drug community, and obtain the drug-target networks by connecting the drugs in the same community through their targets. The connected drug-target networks provide us a useful way to drug discovery.

| Parameters | Values |
|------------|--------|
| $K$        | 150    |
| $\beta$    | 1      |
| $a$        | 450    |
| $b$        |        |
| $\epsilon$ | $10^{-5}$ |

3.1 Quality evaluation of predicted drug community

To evaluate the proposed approach, we compared it with three widely used network partition approaches: Markov cluster (MCL; Enright et al., 2002), affinity propagation (AP; Frey and Dueck, 2007) and ClusterONE (Nepusz et al., 2012). To evaluate the quality of the predicted drug community, the following three metrics are defined. The average enrichment score, $S_N$, is defined as $S_N = \frac{\sum_{i=1}^{N} s_{Ni}}{N}$, where $s_i = \frac{p_i}{P_i}$, $P_i$ is the number of drug pairs sharing targets in the $i$-th drug community, $N_i$ is the number of drugs of that drug community and $N$ is the number of drug community. Higher $S_N$ value indicates more drugs in the same community sharing common targets. The metric, $S_{GR}$, is defined as the relative number of well-connected drug communities $S_R = \frac{|\text{GR}|}{|\text{dGO}|}$, where $|\cdot|$ is the number of elements, and $d$ is a given threshold. The geometric accuracy metric, $S_{GA}$, is defined as the balance of the two metrics: $S_{GA} = \sqrt{S_N \cdot S_R}$. The comparison result shown in Figure 4 shows that the proposed method outperforms the other three methods in terms of all three metrics with $d = 0.6$.

3.2 Drug combination prediction evaluation

To evaluate the prediction capacity of DrugComboRanker, we apply it on the lung adenocarcinoma and endocrine receptor (ER) positive breast cancer. The 107 lung adenocarcinoma patient samples (58 lung tumor tissues and 49 normal lung tissues, GDS3257) were collected to reconstruct the lung adenocarcinoma-specific signaling network, as shown in Figure 3. This signaling network consists of 905 genes and 11 282 interactions. Based on the signaling network, we predict effective drugs that have synergistic effects combined with three drugs (Gefitinib, Paclitaxel and LY-294002) given by our clinicians. Level 3 gene expression data of 222 ER positive breast invasive carcinoma (BRCA) samples and 61 normal breast samples were collected from TCGA. The synergistic drug combination with two FDA-approved drugs, Tamoxifen and Letrozole, are further
predicted for ER-positive BRCA based on the reconstructed ER-positive breast cancer signaling network.

To validate the prediction results, we searched the literature evidence of the top 50 ranked combinations. Tables 2–4 show the literature evidence of the effective drug combinations in our top 50 lists for lung adenocarcinoma. Surprisingly, 19 different drug combinations have been reported to be synergistic combining with Gefitinib, Paclitaxel and LY-294002 in non–small-cell lung cancer. Tables 5 and 6 show the literature evidence of effective drug combinations in the top 50 lists for ER-positive breast cancer. Also 14 different drug combinations have been reported to be synergistic combining with Tamoxifen and Letrozole. These results show the strong drug combination prediction capacity of DrugComboRanker.

To make the evaluation fair and sound, we further compared the predicted results with CDA, and a random combination method (RCM), which randomly picks up 50 drugs from the available drug lists to combine with the designated drugs. As for CDA, we picked the top 50 drug combinations with the designated drugs. Figures 5 and 6 show the comparison results on the lung adenocarcinoma and ER-positive breast cancer in terms of literature supports of those top-ranked 50 drug combinations, respectively (The numbers on the bars of RCM are standard deviations). For RCM, we repeated the random selection 100 times; for each simulation, we checked the literature evidence (In total, we checked all the 5000 random combinations). As can be seen, the proposed approach outperforms the CDA and random selection significantly.

The predicted drug targets in the disease-specific network could indicate the molecular mechanism of synergistic drug combinations. Here, we map the responsive genes of Gefitinib, Paclitaxel, Vorinostat, LY-294002 and Quercetin to the lung adenocarcinoma-specific signaling network to capture the distinct synergistic responses induced by three agent combinations, Gefitinib and Paclitaxel, LY-294002 and Quercetin, Gefitinib and Vorinostat. As shown in Figure 7, Gefitinib and Paclitaxel combinations can affect the EGFR signaling pathway (endothelial cell proliferation), TP53 signaling pathways, as well as

### Table 2. Synergistic alternative drugs combining with Gefitinib

| Drug combination (Community number) | Synergistic score | Rank | Literature evidence |
|-------------------------------------|------------------|------|---------------------|
| Gefitinib/Paclitaxel (63/55)        | 2.905            | 2    | PMID:19596955       |
| Gefitinib/Celecoxib (63/48)         | 2.804            | 3    | PMID:18379355       |
| Gefitinib/Gefitinib (63/102)        | 2.765            | 4    | PMID:22160570       |
| Gefitinib/Fulvestrant (63/55)       | 2.529            | 9    | PMID:24268810       |
| Gefitinib/Vorinostat (63/66)        | 2.468            | 11   | PMID:16713012       |
| Gefitinib/Vorinostat (63/102)       | 2.464            | 12   | PMID:21271222       |
| Gefitinib/Lovastatin (63/34)        | 2.102            | 27   | PMID:19760159       |
| Gefitinib/MS-275(63/102)            | 2.023            | 32   | PMID:168386327      |

### Table 3. Synergistic alternative drugs combining with LY-294002

| Drug combination (Community number) | Synergistic score | Rank | Literature evidence |
|-------------------------------------|------------------|------|---------------------|
| LY-294002/Quercetin (55/140)        | 2.987            | 2    | PMID:14688022       |
| LY-294002/SB-202190(55/5)           | 2.936            | 3    | PMID:16115952       |
| LY-294002/Rapamycin(55/54)          | 2.448            | 33   | PMID:22355375       |

### Table 4. Synergistic alternative drugs combining with Paclitaxel

| Drug combination (Community number) | Synergistic score | Rank | Literature evidence |
|-------------------------------------|------------------|------|---------------------|
| Paclitaxel/Irinotecan(55/66)        | 3.096            | 2    | PMID:16251879       |
| Paclitaxel/Gefitinib (63/63)        | 2.905            | 8    | PMID:19596955       |
| Paclitaxel/Vorinostat(55/102)       | 2.901            | 9    | PMID:19621389       |
| Paclitaxel/Rapamycin(55/54)         | 2.888            | 11   | PMID:22986668       |
| Paclitaxel/Vinblastine(55/137)      | 2.681            | 17   | PMID:16193638       |
| Paclitaxel/Ifosfamide(55/90)        | 2.377            | 25   | PMID:8740798        |
| Paclitaxel/SuraminSodium (55/3)     | 2.189            | 37   | PMID:11507065       |
| Paclitaxel/Resveratrol(55/48)       | 2.060            | 49   | PMID:14666716       |

### Table 5. Synergistic alternative drugs combining with Tamoxifen

| Drug combination (Community number) | Synergistic score | Rank | Literature evidence |
|-------------------------------------|------------------|------|---------------------|
| Tamoxifen/Tretinoin(49/55)          | 2.913            | 8    | PMID:9533531        |
| Tamoxifen/Gefitinib (49/102)        | 2.910            | 9    | PMID:17295255       |
| Tamoxifen/Quercetin(49/140)         | 2.831            | 12   | PMID:20804812       |
| Tamoxifen/Verapamil(49/66)          | 2.776            | 19   | PMID:8185686        |
| Tamoxifen/Mifepristone(49/66)       | 2.721            | 21   | PMID:9879777        |
| Tamoxifen/Celecoxib(49/48)          | 2.437            | 29   | PMID:23731702       |
| Tamoxifen/Resveratrol(49/48)        | 2.416            | 31   | PMID:23896596       |
| Tamoxifen/Letrozole(49/92)          | 2.299            | 36   | PMID:15026471       |
| Tamoxifen/Exemestane(49/51)         | 2.136            | 42   | PMID:21247627       |
| Tamoxifen/LY-294002(49/55)         | 2.035            | 46   | PMID:12479367       |

### Table 6. Synergistic alternative drugs combining with Letrozole

| Drug combination (Community number) | Synergistic score | Rank | Literature evidence |
|-------------------------------------|------------------|------|---------------------|
| Letrozole/Fulvestrant(92/55)        | 2.457            | 8    | PMID:15958593       |
| Letrozole/Tamoxifen(92/49)          | 2.300            | 10   | PMID:15026471       |
| Letrozole/Metformin(92/56)          | 2.148            | 15   | NCT01859367         |
| Letrozole/Celecoxib(92/48)          | 2.128            | 16   | PMID:19254941       |
| Letrozole/Iatininib(92/55)          | 1.765            | 40   | NCT00338728         |
Fig. 5. The comparison results of DrugComboRanker, CDA and RCM in terms of the number of literature supports of the top-ranked 50 drug combinations of lung adenocarcinoma with designated drugs, Gefitinib, Paclitaxel and LY-294002.

Fig. 6. The comparison results of DrugComboRanker, CDA and RCM in terms of the number of literature supports of the top-ranked 50 drug combinations of ER-positive breast cancer with designated drugs, Tamoxifen and Letrozole.

Fig. 7. Drug targets mapped on the disease signaling network. Red and green are the drug targets of Gefitinib and Paclitaxel, respectively.
Fig. 8. Drug targets mapped on the disease signaling network. Red and green are the drug targets of LY-294002 and Quercetin, respectively; blue nodes are the weak affected targets of both drugs.

Fig. 9. Drug targets mapped on the disease signaling network. Red and green are the drug targets of Gefitinib and Vorinostat, respectively.
biological processes, such as cell cycle, apoptosis and the hub genes, i.e. EGFR, TP53, SRC, FOS, JUN. Distinctly the LY-294002 and Quercetin combination affects the alternative EGFR, PI3K-AKT and JAK-STAT3 pathways, as can be seen in Figure 8. In addition, the drug combination, Gefitinib and Celecoxib, targets the EGFR and COX-2 signaling pathways, respectively. The Gefitinib and Celecoxib have distinct transcriptional responses that indicate EGFR and COX-2 signaling pathways are complementary, and have cross talks. Another example is Gefitinib and Vorinostat combination. Vorinostat is a Histone deacetylases inhibitor, as shown in Figure 9, which interacts with CTNNB1, and CTNNB1 interacts with E-cadherin, ERBB2 and EGFR, whereas Gefitinib targets on EGFR. Thus, this combination forms a double inhibition on growth factors.

4 DISCUSSION AND CONCLUSION

Drug combinations can provide an optimal cancer therapeutic treatment by overcoming acquired drug resistance. Clinicians usually choose drug combinations manually based on their experience and expertise. However, this expert-picking-based method cannot be scaled up for discovering drug combinations from a large set of drugs. The availability of genomic profiles of drugs and patients are informative for uncovering the regulatory signaling networks of cancers and the mechanisms of action of drugs. However, the computational tools are the bottleneck to interpret and convert the big genomics data into discovery of drug combinations. In this study, we propose a novel computational tool, DrugComboRanker, to prioritize drug combinations for specific cancers by selecting combinations targeting the alternative and complementary signaling modules of disease. We first predict the functional targets of drugs via the network community analysis of drug functional networks based on their genomic profiles. We then define the synergistic score of drug combinations that block key disease genes and their potential cross talks. The new method can also provide insights into mechanism of actions of drug combinations by mapping the predicted drug targets on the disease signaling network. In our case studies on lung adenocarcinoma and ER-positive breast cancer cases, a set of top-ranked drug combinations, with similar and distinct mechanisms of action, top-ranked in our prediction list have been reported to be effective in anticancer treatment. Moreover, a set of novel drug combinations are also identified with a better anticancer treatment outcome. In conclusion, the DrugComboRanker has the potential to be a widely used software package for predicting new disease-specific drug combinations.

ACKNOWLEDGEMENTS

The author would like to thank the helpful discussion with Drs Fredrik Barrenäs, Rui Zhu and Wen Bu.

Funding: CPRIT (RP110532) and NCI (U54 CA149196). Ting Tsung and Wei Fong Chao Foundation and John S Dunn Foundation

Conflict of Interest: none declared.

REFERENCES

Awan, A. et al. (2007) Regulatory network motifs and hotspots of cancer genes in a mammalian cellular signalling network. IET Syst. Biol., 1, 292-297.

Barrenäs, F. et al. (2012) Highly interconnected genes in disease-specific networks are enriched for disease-associated polymorphisms. Genome Biol., 13, 1-9.

Brandes, U. and Fleischer, D. (2005) centrality measures based on current flow. In: Dietker, V. and Durand B. (eds) STACS 2005. Springer, Berlin, Heidelberg, pp. 533-544.

Chuang, H.-Y. et al. (2007) Network-based classification of breast cancer metastasis. Mol. Syst. Biol., 3, 140.

Couto, F.-M. et al. (2007) Measuring semantic similarity between Gene Ontology terms. Data Knowl. Eng., 61, 137–152.

Cui, Q. et al. (2007) A map of human cancer signaling. Mol. Syst. Biol., 3, 152.

Diaconis, P. and Graham, R. L. (1977) Spearman’s Footrule as a measure of disarray. J. R. Stat. Soc. B Methodol., 39, 262–268.

Enright, A. J. et al. (2002) An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res., 30, 1575–1584.

Frey, B. J. and Dueck, D. (2007) Clustering by passing messages between data points. Science, 315, 972–976.

Ideker, T. et al. (2002) Discovering regulatory and signalling circuits in molecular interaction networks. Bioinformatics, 18, S23-S240.

Iorio, F. et al. (2010) Discovery of drug mode of action and drug repositioning from transcriptional responses. Proc. Natl Acad. Sci. USA, 107, 14621–14626.

Jensen, L. J. et al. (2006) Literature mining for the biologist: from information retrieval to biological discovery. Nat. Rev. Genet., 7, 119–129.

Johannsen, S. et al. (2013) Sevoflurane is less sensitive than halothane for in vitro detection of malignant hyperthermia susceptibility. Acta Anaesthesiol. Scand., 57, 1161–1166.

Kuhn, M. et al. (2012) STITCH 3: zoning in on protein-chemical interactions. Nucleic Acids Res., 40, D876–880.

Lamb, J. et al. (2006) The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. Science, 313, 1929–1935.

Lee, J.-H. et al. (2012) CDA: combinatorial drug discovery using transcriptional response modules. PLoS One, 7, e42573.

Li, J. et al. (2012) The human phosphotyrosine signaling network: evolution and hotspots of hijacking in cancer. Genome Res., 22, 1222–1230.

Lin, S. (2010) Space oriented rank-based data integration. Stat. Appl. Genet. Mol. Biol., 9, Article20.

Martin, Y. C. et al. (2002) Do structurally similar molecules have similar biological activity? J. Med. Chem., 45, 4350-4358.

Nazarian, R. et al. (2010) Melanomas acquire resistance to B-RAF(V600E) inhibition by RTK or N-RAS upregulation. Nature, 468, 973–977.

Nepusz, T. et al. (2012) Detecting overlapping protein complexes in protein-protein interaction networks. Nat. Methods, 9, 471-472.

Newman, R. H. et al. (2013) Construction of human activity-based phosphorylation networks. Mol. Syst. Biol., 9, 655.

Page, L. et al. (1999) The PageRank Citation Ranking: Bringing Order to the Web. Stanford InfoLab, Stanford University, Stanford, CA.

Šćiric, J. et al. (2006) Extraction of regulatory gene/protein networks from Medline. Bioinformatics, 22, 645–650.

Stark, C. et al. (2006) BioGRID: a general repository for interaction datasets. Nucleic Acids Res., 34, D535–D539.

Subramanian, A. et al. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl Acad. Sci. USA, 102, 15545–15550.

Tan, Y. F. and Fovetone, C. (2013) Automatic relevance determination in nonnegative matrix factorization with the beta-divergence. IEEE Trans. Pattern Anal., 35, 1592–1605.

Zhao, X.-M. et al. (2011) Prediction of drug combinations by integrating molecular and pharmacological data. PLoS Comput. Biol., 7, e1002323.

Zhou, T. et al. (2010) Solving the apparent diversity-accuracy dilemma of recombinant systems. Proc. Natl Acad. Sci. USA, 107, 4511-4515.

Zinner, R. G. et al. (2009) Algorithmic guided screening of drug combinations of arbitrary size for activity against cancer cells. Mol. Cancer Ther., 8, 521–532.