Bioinformatics Advance Access published August 18, 2016

Systems biology

Predicting synergistic effects between compounds through their structural similarity and effects on transcriptomes

Yiyi Liu[1] and Hongyu Zhao[1,2,*]

[1]Department of Biostatistics, Yale School of Public Health, New Haven, CT,

[2]Program of Computational Biology and Bioinformatics, Yale University, New Haven, CT, USA.

[*]To whom correspondence should be addressed.

Associate Editor: Dr. Jonathan Wren

Abstract

Motivation: Combinatorial therapies have been under intensive research for cancer treatment. However, due to the large number of possible combinations among candidate compounds, exhaustive screening is prohibitive. Hence, it is important to develop computational tools that can predict compound combination effects, prioritize combinations and limit the search space to facilitate and accelerate the development of combinatorial therapies.

Results: In this manuscript we consider the NCI-DREAM Drug Synergy Prediction Challenge dataset to identify features informative about combination effects. Through systematic exploration of differential expression profiles after single compound treatments and comparison of molecular structures of compounds, we found that synergistic levels of combinations are statistically significantly associated with compounds' dissimilarity in structure and similarity in induced gene expression changes. These two types of features offer complementary information in predicting experimentally measured combination effects of compound pairs. Our findings offer insights on the mechanisms underlying different combination effects and may help prioritize promising combinations in the very large search space.

Availability: The R code for the analysis is available on https://github.com/YiyiLiu1/DrugCombination.

Contact: hongyu.zhao@yale.edu

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Designing effective therapies for cancer treatment is an important yet challenging task. Although monotherapies are commonly used to treat cancer (Joensuu, et al., 2001; Vogel, et al., 2002), they suffer from many problems, such as acquired resistance and poor safety (LoRusso, et al., 2012). Combinatorial therapies, by utilizing compounds impacting multiple biological processes simultaneously, have great potentials to overcome these problems and are attracting growing interest in drug development (Al-Lazikani, et al., 2012; Jia, et al., 2009). Indeed, there are many effective combinations used in practice. For example, the combination of Lapatinib and Capecitabine can achieve improved efficacy in breast cancer treatment (Geyer, et al., 2006). In this study we focus on two-way combinations because most studies and available data to date are on two-way combinations and yet all existing methods are still limited in their performance in predicting the effectiveness of two-way combinations. Generally, when two compounds act together, their combination effect can be categorized into three main types: additive, if the effect of combination is equivalent to the sum of the effects of two compounds acting individually; synergistic, if the combination effect is greater than additive; and antagonistic, if the combination effect is less than additive. Typically compound combination effects are inferred through cell culture experiments, where cell viabilities under treatment with compounds in combination and treatment with compounds individually are measured and compared. However, since the number of possible combinations grows rapidly with the number of compounds under consideration, exhaustive experimental screening of all these combinations is prohibitively costly. Therefore, it is very useful to develop computational methods that can prioritize different combinations in order to

reduce the search space for screening experiments. There are inspiring attempts made towards this goal (Guimera and Sales-Pardo, 2013; Huang, et al., 2013; Huang, et al., 2014; Li, et al., 2015; Pang, et al., 2014; Zhao, et al., 2011). However, while it is known combination effects can be cell-line specific (Held, et al., 2013), most existing computational methods were trained using only the specific types of combination effects (synergistic/additive/antagonistic) collected from literature review or databases, without considering the cell lines they were tested on or other experimental contexts. Besides, many of these methods rely heavily on prior knowledge of drug combination mechanisms, which is far from complete and accurate.

In this paper, we aim to identify features informative on combination effects of compound pairs using the dataset from the NCI-DREAM Drug Synergy Prediction Challenge (Bansal, et al., 2014). This DREAM challenge measured gene expression profiles of OCL-LY3 diffuse large B-cell lymphoma cells pre- and post- 14 single compound treatments, and experimentally evaluated the combination effects of 91 pairwise combinations of these 14 compounds. In the original challenge, participants proposed different methods to predict combination effects using gene expression profiles, but no unanimous conclusion could be drawn on the relation between synergistic effects and transcriptome changes induced by single compounds (Bansal, et al., 2014). Although a few participants considered compound chemical/molecular features such as structure, the relevance of these features to combination effects was not demonstrated. In this paper, we adopt the structural similarity of compounds defined in PubChem (Kim, et al., 2016) and consider four similarity measures for gene expression changes resulting from single compound treatments. We systematically explore whether the structural similarity and expression similarity is associated with combination effects using PC-index and resampled Spearman correlation (Bansal, et al., 2014). Because we are interested in prioritizing compound pairs likely having synergistic effects, we also evaluate the area under the receiver operation characteristic curve (AUC) for synergistic combination prediction using structural similarity and expression similarity individually and jointly. For the NCI-DREAM challenge data set, we find that synergistic effect prediction using gene expression similarity defined with one of the four measures outperforms the best method reported in the original challenge, and that prediction using structural similarity also outperforms most of the methods proposed for the original challenge (Bansal, et al., 2014). Moreover, combining gene expression and structure information together further improves predictive power for synergistic effects. Our findings in this paper may lead to better computational methods to prioritize compound combinations and facilitate future drug development.

2 Methods

2.1 Data preprocessing

We analyzed the data distributed for the NCI-DREAM Drug Synergy Prediction Challenge (Bansal, et al., 2014; Goswami, et al., 2015; Yang, et al., 2015). In this project, the OCL-LY3 cell line was perturbed by 14 single compounds, each with 2 concentrations (IC_{20}'s of 24h and 48h), and gene expression profiles of untreated samples, DMSO-treated (DMSO was used as a control media) and single compound-treated samples were generated at three time points (6h, 12h and 24h after treatment). All expression profiles were measured using the Human Genome U219 96-Array plate (Affymetrix), in triplicate except that DMSO-treated ones were in octuplicate. Quality-controlled and RMA-normalized (Irizarry, et al., 2003) data (in log2 scale) were provided by the Challenge. All 91 pairwise combinations of the 14 compounds (14×13/2) were tested with each compound concentrated at its IC_{20} of 60h. Combination effects of these compound pairs were assessed using the Bliss independence model (Bliss, 1939). Mean Excess over Bliss (EOB) values (estimated from five replicates) were provided as a measurement of the synergy levels and standard errors (SE) were also given to quantify the uncertainties. We used these experimentally measured combination effects as "gold standard" in our analysis.

We first applied RUVr (Risso, et al., 2014) to remove potential batch effects in gene expression measurements. We considered compounds, concentrations and time points as treatment status ("wanted variation") and varied the number of hidden factors ("unwanted variation") k from 1 to 40. Based on the relative log expression (RLE) plots (Figures S1 and S2), we set $k = 26$ in our analysis. We also reported results with other values of k in supplementary information (Figures S3-S10 and Tables S1-S5), which suggest that our results are quite robust to the specific choice of k.

2.2 Structural similarity measure

We used substructure key-based 2D Tanimoto similarity score from the PubChem database (Kim, et al., 2016) to quantify the structural similarity between two compounds. PubChem generates a binary fingerprint (an ordered list of binary bits) to represent the presence or absence of specific chemical substructure for each compound. It then defines the similarity score between two compounds as

$$Tanimoto\ score = \frac{AB}{A + B - AB},$$

where AB is the count of bits shared by the two compounds, and A and B are the counts of bits in the two compounds, respectively.

It can be seen from the above formula that the Tanimoto score is always between 0 and 1. According to PubChem, a Tanimoto score of 0.68 or greater is considered statistically significant at the 95% significance level (Kim, et al., 2016).

2.3 Gene expression similarity measures

We used the average of three/eight replicates to represent gene expression levels under each experimental condition. We then calculated the fold changes between treatments (single compounds) and control (DMSO). The following metrics were adopted to evaluate the global differential expression similarities between two drugs.

- Direction-based

This metric considers the concordance of the direction of gene expression changes (up or down) following single-compound treatments. It was motivated by the second best performing method in the original challenge (Goswami, et al., 2015). However, unlike what was done there, we did not assume a common core gene set for all compounds since different compounds may impact different genes and pathways. Instead, we selected out genes with large expression changes for each compound and measured the concordance of two compounds on their corresponding "signature" genes. Specifically, we discretized expression changes into three levels: 0 for probes with fold changes within a certain range around 1 and thus viewed as unaffected by a compound; -1 and 1 for probes with fold changes outside the range and considered down-regulated and up-regulated, respectively. We then define the expression similarity score between compound i and compound j as

$$\frac{\sum_{l=1}^{n} fc_{i,l} \times fc_{j,l}}{n},$$

where n is the total number of probes (48,789 in this dataset) and $fc_{i,l}$ is the status (-1, 0 or 1) for probe l treated with compound i.

- GSEA-based

This metric was proposed in the context of constructing drug similarity network with transcriptome data (Iorio, et al., 2010; Iorio, et al., 2009). The original method (Iorio, et al., 2010; Iorio, et al., 2009) first ranks probes from the most up-regulated to the most down-regulated in each of the six datasets measured at three time points and two concentrations, and then merges the six lists using pair-wise Spearman's Footrule, Borda Merging Method and Kruskal Algorithm (Diaconis, 1977; Lin, 2010) into a unified ranked list hierarchically. Then for compounds i and j, it calculates four Gene Set Enrichment Analysis (GSEA) (Subramanian, et al., 2005) scores: ES_i^{jp} and ES_i^{jq}, the enrichments of compound j's top p probes and bottom q probes in compound i's expression list; ES_j^{ip} and ES_j^{iq}, the enrichments of compound i's top p probes and bottom q probes in compound i's expression list. A distance between i and j is defined as:

$$d_{ij} = 1 - \frac{\left(\frac{ES_i^{jp} - ES_i^{jq}}{2} + \frac{ES_j^{ip} - ES_j^{iq}}{2}\right)}{2}.$$

Since the values of ES's are between -1 and 1, d_{ij} is between 0 and 2. In our analyses, we define the corresponding similarity score as:

$$1 - \frac{d_{ij}}{2},$$

which is between 0 to 1. We also calculated d_{ij} without merging the six lists for each compound.

- Pearson correlation-based

Under each treatment condition, we ranked all the probes according to their absolute log2-transformed intensity differences, and then selected top probes as signatures. We calculated the Pearson correlation between two compounds' expression profiles on all 48,789 probes as well as on the union of their signature gene sets.

- Spearman correlation-based

We also calculated the Spearman correlation between two compounds' expression profiles on all 48,789 probes as well as on the union of their signature gene sets.

2.4 Evaluation of concordance between similarity-based scores and synergistic levels of compound pairs

We used probabilistic concordance index (PC-index) and resampled Spearman correlation to quantify the concordance between the similarity-based scores and synergistic levels of compound pairs (Bansal, et al., 2014).

2.4.1 PC-index

Suppose that there are a total of N compound pairs. Let the similarity-based scores for compound pair n be s_n, and the experimentally measured EOB and its standard error be eob_n and se_n, then for two compound pairs m and n, we first compute

$$sp_{mn} = \begin{cases} \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{eob_m - eob_n}{\sqrt{se_m^2 + se_n^2}}\right)\right), & if\ s_m > s_n \\ \frac{1}{2}\left(1 - \mathrm{erf}\left(\frac{eob_m - eob_n}{\sqrt{se_m^2 + se_n^2}}\right)\right), & if\ s_m < s_n \\ \frac{1}{2}, & otherwise \end{cases}$$

where

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x exp(-t^2)\ dt$$

PC-index is defined as

$$\frac{2}{N \times (N-1)} \sum_{\substack{m=1,2,\cdots,N-1 \\ n=m+1,2,\cdots,N}} sp_{mn}.$$

PC-index takes experimental errors into account when evaluating the concordance between the predictions and the "gold standard". It is symmetric around 0.5, i.e., if a prediction has pc-index pc, then flipping it over (exactly opposite order) gives pc-index $1 - pc$. The largest value (PCmax) is reached when the predictions are entirely concordant with the "gold standard". Due to noise terms, PCmax will not be 1 generally. In this study,

$$PC_{max} \approx 0.901.$$

2.4.2 Resampled Spearman correlation

Similar to PC-index, resampled Spearman correlation accounts for uncertainties of the experiment. It assumes that the experimental measurement of mean EOB is noisy and follows a normal distribution, $N(\mu, \sigma^2)$, with μ and σ equal to the mean EOB and standard error obtained from the experiment. For each compound pair n, we randomly sample a new eob_n^* from $N(\mu_n, \sigma_n^2)$, and then calculated the spearman correlation between $(p_1, p_2, \cdots, p_N)$ and $(eob_1^*, eob_2^*, \cdots, eob_N^*)$. We repeat the process for 10,000 times and use the mean of the 10,000 Spearman correlations as a final resampled Spearman correlation score rss.

2.4.3 Statistical significance estimation

We simulated 10,000 independent random predictions (10,000 random permutations of the 91 pairs) and calculated their PC-indices and resampled Spearman correlations. We use these scores from random predictions as the empirical null distributions for PC-index and resampled Spearman correlation. Then we estimated the p-value of $PC - index$ and rss for each similarity-based prediction as

$$P - value_{PC-index} = \frac{\#(PC - index_{null} \geq PC - index)}{10{,}000},$$

$$P - value_{rss} = \frac{\#(rss_{null} \geq rss)}{10{,}000}.$$

2.5 Synergistic combination prediction

We treated compound pairs with experimentally measured synergistic combination as positive and all others as negative, and adopted logistic regression models to predict synergistic combinations using gene expression similarity and structural similarity, individually and jointly. We performed 100 rounds of 3-fold cross validation. In each round, we estimated the cross-validation AUC using trapezoidal method to integrate the ROC curve. We reported the mean AUC over 100 rounds for each classifier as an evaluation of its performance. To assess the statistical significance of improvement brought by structural information compared to using gene expression only, we randomly permutated the structural similarity scores of the 91 pairs 1,000 times, and for each permutation, we calculated the mean AUC (over 100 rounds of cross validation) of the classifier combining the permutated structural similarity score and the original gene expression similarity score. We calculated the p-value of AUC improvement as

$$P - value_{improvement\ by\ structure} = \frac{\#(AUC_{expression+permutated\ structure} > AUC_{expression+structure})}{1{,}000}.$$

Similarly, we calculated the p–value of AUC improvement by adding gene expression information as

$$P - value_{improvement\ by\ expression} = \frac{\#(AUC_{structure+permutated\ expression} > AUC_{expression+structure})}{1{,}000}.$$

3 Results

We evaluated the structural similarity of the 14 compounds and also compared their effects on the transcriptome of the OCL-LY3 cell line. Following the performance evaluation methods of the original DREAM challenge (Bansal, et al., 2014), we used PC-index and resampled Spearman correlation coefficient to measure the associations between these similarity scores and the synergistic levels. Additionally, we evaluated these scores' ability to predict synergistic combinations using AUC from cross validation.

3.1 Structural similarity negatively correlates with synergism

A negative correlation can be observed between EOB for two compounds against their structural similarity scores (Figure 1). The PC-index between EOB and the negative structural similarity score was 0.586 (p-value 0.0047), and the resampled Spearman correlation between them was 0.251 (p-value 0.0041). This result outperforms most of the methods in the original challenge (Table S6). It indicates that synergistic effect is significantly negatively correlated with structural similarity.

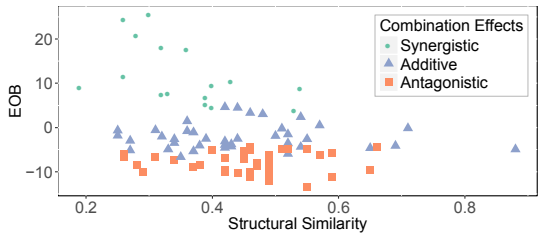

Fig. 1. EOB against structural similarity of compounds. Activities of compound pairs are discretized into three states as defined in the original Challenge: synergistic (green circle), additive (blue triangle) and antagonistic (red square).

In addition, there is an interesting pattern behind this overall negative correlation. In this dataset, compound pairs with EOB > 0 and EOB/SE > 2 were classified as having synergistic effects; similarly, those with EOB < 0 and EOB/SE < -2 were defined to be antagonistic; and the rest of the combinations were classified as additive (Bansal, et al., 2014). We observed that none of the compound pairs with large structural similarity scores (e.g., > 0.6) have synergistic effects, while the combination effects of the compound pairs with small structural similarity scores spread across all the three classes. We will discuss possible mechanisms for this phenomenon in the Discussion and Conclusion section.

3.2 Gene expression similarity positively correlates with synergism

We considered four measures for the similarity of gene expression changes brought by two compounds individually and all the similarity scores showed positive correlations with EOB.

We defined the first similarity score based on the directions (up or down) of gene expression changes. We set different thresholds on fold changes to define signatures and calculated the PC-indices and resampled Spearman correlations between EOB and the direction-based similarity measurements (Figure 2 (a-b)). In all situations we considered, the PC-indices were greater than 0.5 and the resampled Spearman correlations were greater than 0, which indicate positive correlations between synergistic level and gene expression similarity. Taking all genes into account (green bars in Figure 2), only weak correlations could be observed; by excluding genes with small changes, signals in most of the six datasets (measured at three time points and two concentrations) became strong. The largest PC-index and resampled Spearman correlation were 0.618 (p-value 0.0003) and 0.338 (p-value 0.0002), respectively.

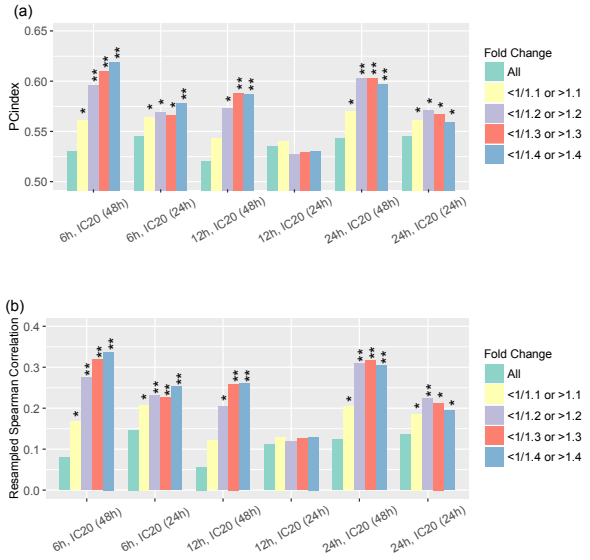

Fig. 2. (a) PC-indices of direction-based gene expression similarity scores. (b) Resampled Spearman correlations of direction-based gene expression similarity scores. When calculating the similarity scores, we considered all probes as well as probes with fold changes beyond certain ranges only (shown in different colors). *P-value $<$ 0.05; ** P-value $<$ 0.01.

The second gene expression similarity score we considered was based on GSEA. We set different cutoffs to define top and bottom signature genes and the resulting PC-indices and resampled Spearman correlations between these expression similarities and EOB were shown in Figure 3 (a-b). Again, we could observe positive correlations between EOB and these expression similarities in all cases. Here the largest PC-index and resampled Spearman correlation we got were 0.574 (p-value 0.0125) and 0.213 (p-value 0.0112), respectively.

The third and fourth expression similarities we considered were based on the more conventional Pearson and Spearman correlations. Similar to what we did earlier, besides considering all the probes, we also selected strong signals (probes with the largest expression changes under the treatment of each single compound) and calculated the correlations only on these probes. We presented the PC-indices and resampled Spearman correlations between EOB and these two correlation-based scores in Figures 4 (a-b) and 5 (a-b), respectively. These two expression similarity measures also had positive correlations with synergistic levels. The largest PC-index and the largest resampled Spearman correlation we got for the Pearson correlation-based similarity were 0.582 (p-value 0.0068) and 0.245 (p-value 0.0046), respectively. The largest PC-index and resampled Spearman correlation we got for Spearman correlation-based similarity were 0.571 (p-value 0.0156) and 0.206 (p-value 0.0147), respectively.

Similar to structural similarity, we observed (Figure S11) that it is less likely for the compound pairs with relatively small gene expression similarity scores to have synergistic effects, while for the compound pairs with relatively large gene expression similarity scores, their combination effects vary case by case.

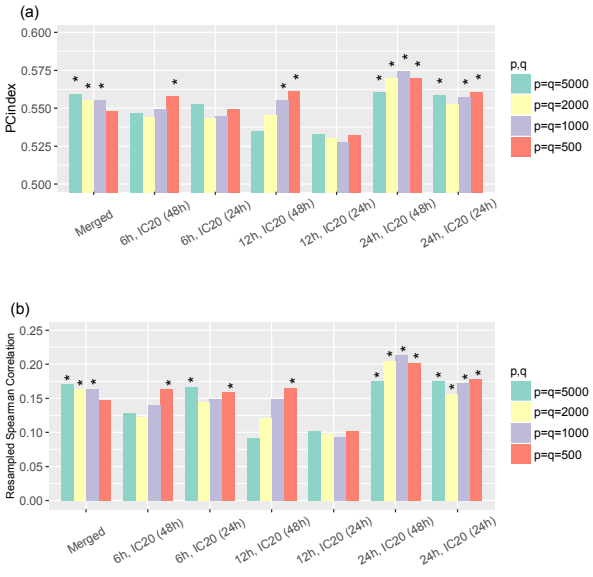

Fig. 3. (a) PC-indices of GSEA-based gene expression similarity scores. (b) Resampled Spearman correlations of GSEA-based gene expression similarity scores. When calculating the similarity scores, we considered different numbers of top (p) and bottom (q) signatures (shown in different colors); we included the results for a merged data set and six unmerged data sets separately. *P-value < 0.05; *P-value < 0.01.

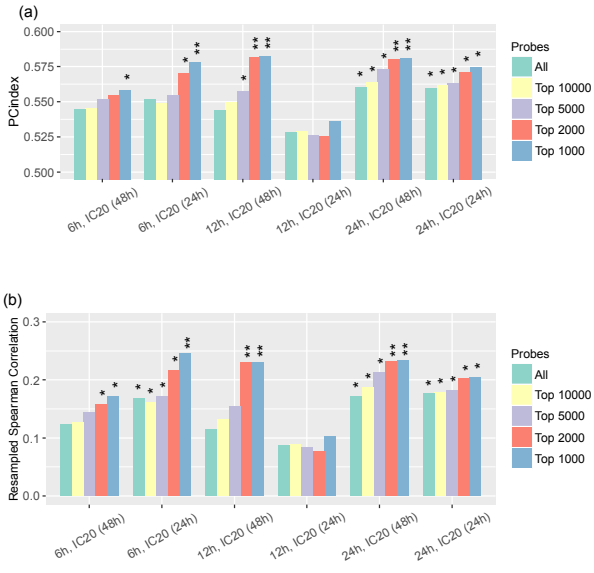

Fig. 4. (a) PC-indices of Pearson correlation-based gene expression similarity scores. (b) Resampled Spearman correlations of Pearson correlation-based gene expression similarity scores. When calculating the similarity scores, we considered all probes as well as probes with largest expression changes only (shown in different colors). *P-value < 0.05; *P-value < 0.01.

In addition, we also note the following two observations. First, according to our results, all six datasets have certain information that could be utilized, and no dosage or exposure time is "optimal" in terms of transcriptome concordance with experimentally measured synergistic levels. Second, we found that signals can usually be strengthened when we focus on a subset of signature genes influenced most by the perturbations rather than considering all of them together. This is reasonable as many genes may not be impacted by individual drug perturbations and introduce noises.

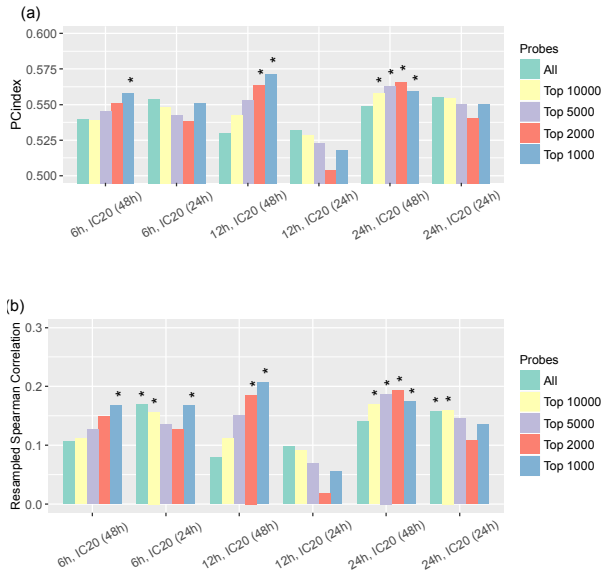

Fig. 5. (a) PC-indices of Spearman correlation-based gene expression similarity scores. (b) Resampled Spearman correlations of Spearman correlation-based gene expression similarity scores. When calculating the similarity scores, we considered all probes as well as probes with largest expression changes only (shown in different colors). *P-value < 0.05; *P-value < 0.01.

3.3 Structural similarity and gene expression similarity provide complementary information on synergistic combination prediction

Since compound pairs with synergistic effects are most worthy of further investigation, we trained logistic regression models to predict synergistic combinations using structural similarity score, each of the four gene expression similarity scores, individually and jointly. The AUCs estimated from 100 rounds 3-fold cross-validations are listed in Table 1 (a-e).

It can be seen that most classifiers achieved satisfactory results in identifying synergistic combinations. Moreover, using structural similarity and gene expression similarity together does lead to greater predictive power. Specifically, for direction-based gene expression similarity scores, combining them together with structural similarity, in all the 30 cases (6 datasets × 5 thresholds), 28 had larger AUCs compared with using expression similarity alone (all with p-values < 0.05), and 23 had larger AUCs compared with using structural similarity alone (22 with p-values < 0.05); for GSEA-based gene expression similarity scores, among the 28 cases (7 datasets × 4 thresholds), combining them together with structural similarity outperformed using expression similarity alone and using structural similarity alone in 28 cases (all with p-values < 0.05) and 19 cases (8 with p-values < 0.05), respectively; for Pearson correlation-based gene expression similarity scores, when using two features, 29 out of 30 cases (6 datasets × 5 thresholds) had larger AUCs than using expression similarity alone (all with p-values < 0.05) and 15 had larger AUCs than using structural similarity alone (7 with p-values < 0.05); for Spearman correlation-based gene expression similarity

scores, AUCs in all the 30 cases (6 datasets × 5 thresholds) were larger when using both features than using expression similarity alone (all with p-values < 0.05), and AUCs in 13 cases were larger than using structural similarity alone (6 with p-values < 0.05). Hence, we concluded for most of these gene expression similarity scores, their predictions for synergistic combinations were significantly improved when incorporating structural information. Especially in the cases where expression similarity scores were less informative (Figures 2-5 and Table 1), the increments were remarkable; as it is not always easy to identify an "optimal" dosage/duration in transcriptome effect measurements, we believe including structural information would benefit synergistic combination prediction greatly. Besides, for most cases where gene expression similarity demonstrated certain power to predict synergistic combinations, combining them and structural similarity together did lead to better performance over using structural similarity alone, too. Additionally, we found that the dataset measured at 6h post-treatment with IC_{20}'s of 24h was most useful here compared to others, a phenomenon not observed when examining the scores' correlations to EOBs (Section 3.2). We investigated the expression similarity scores on this dataset and found that most of them indeed could separate out synergistic pairs from the rest more effectively, as compared to measurements obtained from other datasets, synergistic pairs tend to be more concentrated on the relatively large gene expression similarity parts using these scores; while PC-index and resampled Spearman correlation emphasize more on the global concordance between the similarity score and EOBs, such signal was diluted.

In summary, we observed that both structural similarity and gene expression similarity are informative to predict synergistic effects; and these two types of features provide complementary information.

Table 1. AUC of logistic regression model to predict synergistic combinations

1a. Direction-based expression similarity scores

	6h, IC20 (48h)	6h, IC20 (24h)	12h, IC20 (48h)	12h, IC20 (24h)	24h, IC20 (48h)	24h, IC20 (24h)
All	0.43 [0.70]	0.70 [0.77]	0.47 [0.71]	0.67 [0.76]	0.56 [0.73]	0.56 [0.71]
FC >1.1 or <1/1.1	0.55 [0.71]	0.77 [0.80]	0.43 [0.70]	0.69 [0.77]	0.66 [0.76]	0.68 [0.74]
FC >1.2 or <1/1.2	0.69 [0.75]	0.81 [0.83]	0.62 [0.73]	0.74 [0.78]	0.79 [0.82]	0.78 [0.79]
FC >1.3 or <1/1.3	0.75 [0.77]	0.81 [0.81]	0.74 [0.78]	0.74 [0.79]	0.84 [0.85]	0.79 [0.82]
FC >1.4 or <1/1.4	0.77 [0.77]	0.82 [0.80]	0.76 [0.80]	0.73 [0.78]	0.85 [0.84]	0.79 [0.81]

1b. GSEA-based expression similarity scores

	Merged	6h, IC20 (48h)	6h, IC20 (24h)	12h, IC20 (48h)	12h, IC20 (24h)	24h, IC20 (48h)	24h, IC20 (24h)
p=q=5000	0.65 [0.73]	0.51 [0.72]	0.74 [0.78]	0.46 [0.71]	0.66 [0.76]	0.64 [0.74]	0.65 [0.73]
p=q=2000	0.66 [0.74]	0.52 [0.72]	0.75 [0.79]	0.43 [0.71]	0.65 [0.76]	0.67 [0.75]	0.65 [0.73]
p=q=1000	0.66 [0.74]	0.55 [0.72]	0.76 [0.79]	0.42 [0.71]	0.65 [0.75]	0.68 [0.75]	0.66 [0.74]
p=q=500	0.65 [0.74]	0.57 [0.72]	0.77 [0.79]	0.42 [0.71]	0.65 [0.75]	0.68 [0.75]	0.66 [0.74]

1c. Pearson correlation-based expression similarity scores

	6h, IC20 (48h)	6h, IC20 (24h)	12h, IC20 (48h)	12h, IC20 (24h)	24h, IC20 (48h)	24h, IC20 (24h)
All	0.48 [0.71]	0.74 [0.78]	0.42 [0.71]	0.63 [0.75]	0.62 [0.74]	0.64 [0.73]
Top 10000	0.50 [0.71]	0.75 [0.78]	0.42 [0.71]	0.62 [0.74]	0.64 [0.75]	0.64 [0.72]
Top 5000	0.54 [0.72]	0.77 [0.79]	0.42 [0.71]	0.61 [0.74]	0.67 [0.76]	0.64 [0.73]
Top 2000	0.57 [0.72]	0.79 [0.79]	0.46 [0.71]	0.58 [0.73]	0.68 [0.76]	0.62 [0.73]
Top 1000	0.58 [0.72]	0.81 [0.80]	0.46 [0.71]	0.59 [0.73]	0.69 [0.75]	0.60 [0.72]

1d. Spearman correlation-based expression similarity scores

	6h, IC20 (48h)	6h, IC20 (24h)	12h, IC20 (48h)	12h, IC20 (24h)	24h, IC20 (48h)	24h, IC20 (24h)
All	0.45 [0.71]	0.72 [0.78]	0.45 [0.71]	0.65 [0.76]	0.59 [0.73]	0.59 [0.71]
Top 10000	0.48 [0.71]	0.75 [0.78]	0.42 [0.71]	0.63 [0.75]	0.63 [0.75]	0.63 [0.72]
Top 5000	0.54 [0.72]	0.75 [0.79]	0.43 [0.71]	0.59 [0.74]	0.66 [0.75]	0.62 [0.72]
Top 2000	0.60 [0.73]	0.74 [0.79]	0.46 [0.71]	0.50 [0.73]	0.68 [0.75]	0.51 [0.71]
Top 1000	0.61 [0.73]	0.76 [0.79]	0.49 [0.71]	0.48 [0.72]	0.67 [0.75]	0.45 [0.69]

1e. Structural similarity score

Structure	0.73

1a-1d show AUCs using gene expression similarity only and combining gene expression similarity with structural similarity (in brackets). In 1a, "FC" is short for "Fold Change". 1e shows AUC using structural similarity only.

4 Discussion and Conclusion

Developing combinatorial therapies for cancer treatment has attracted increasing attention due to their great potentials as compared to monotherapies. One major challenge to develop combinatorial therapies is the large search space of possible combinations, therefore computationally predicting combination effects and prioritizing combinations is vitally important. In this paper, we have analyzed the experimentally measured 91 combinations of 14 compounds from the NCI-DREAM Drug Synergy Prediction Challenge to identify features that are informative on combination effects.

We observed that structural similarity is statistically significantly negatively correlated with synergism. There has been previous work utilizing structural similarity to model compound combination effects, based on the assumption that if the combination of compound *A* and compound *B* has a specific effect, compounds having similar structure to compound *A* (*B*) tend to have the same interaction with compound *B* (*A*) (e.g. (Vilar, et al., 2012)). Compared with these work and several methods in the original challenge, we systematically investigated the relation between structural similarity and synergism. We hypothesize that the negative

correlation we observed may be caused by "competitive binding" (Cokol, et al., 2011; Jia, et al., 2009), i.e., when two compounds bind to similar targets they tend to have antagonistic or additive effects due to competing and interfering interactions between them. Specifically, when two compounds have similar structures (i.e. many share substructures), they more likely interact with the same protein sites or perturb the same biological processes, where competition exists, and reduces the chance of synergistic effects. On the other hand, compound pairs with distinct structures (lower structural similarity scores) can interact with different proteins or have varied functions, where there is no strong competition for "resources"; depending on the biological processes they influence and genomic context of the cell line they work on, their combination effects can be synergistic, additive or antagonistic.

We also found a statistically significant positive association between synergism and the similarity of gene expression changes caused by single compounds. This association is robust to the four different similarity measures we considered on all six gene expression profiles. We hypothesize that two compounds leading to different changes on gene expressions may offset each other's effects when applied together, thus are less likely able to "collaborate" to generate synergistic effects. On the other hand, if two compounds lead to similar gene expression changes, they may not have such "neutralizing" problem and hence are more likely to be synergistic; yet, the actual effects in such cases can vary depending on the interactions of different biological processes affected by the two compounds.

In addition, we found both gene expression similarity and structural similarity are predictive of synergistic combinations. More importantly, these two distinct similarity scores provided complementary information on synergistic effect predictions; when utilized together, the performance can be further improved. Besides classifying compound combinations as synergistic/non-synergistic, similarly we also observed complementary informativeness of gene expression and structural similarity in the three-class (synergistic, additive and antagonistic) classification as well (s S3-S5). Our findings in this paper should be useful for prioritizing compound combinations. They also provide insights for us to further investigate the mechanisms behind various combination effects.

There are several directions worth explorations in future studies. First, for compound structural similarity scores, we used 2D information here as it is robust and often generates superior results in activity prediction (Maggiora, et al., 2014). In the future, when more accurate descriptions of 3D compound structures and reliable metrics of 3D structural similarity are available, we can carry out similar analysis based on 3D information. Second, we considered gene expression similarity defined on individual genes in this paper. Since genes are not isolated in biological systems, better measures may be defined by taking gene-gene interactions into account. Finally, in addition to structure and expressions, other types of molecular and biological features may be investigated and prove useful in predicting combination effects.

Acknowledgements

We acknowledge Dr. David Stern of Yale Department of Pathology and Dr. Kuang-Yao Lee of Zhao Lab for their helpful discussions.

Funding

This work was supported by NIH grants GM59507, CA154295, and CA196530.

Conflict of Interest: none declared.

References

Al-Lazikani, B., Banerji, U. and Workman, P. Combinatorial drug therapy for cancer in the post-genomic era. *Nat Biotechnol* 2012;30(7):679-691.

Bansal, M., *et al.* A community computational challenge to predict the activity of pairs of compounds. *Nat Biotechnol* 2014;32(12):1213-1222.

Bliss, C.I. The toxicity of poisons applied jointly. *Ann Appl Biol* 1939;26(3):585-615.

Cokol, M., *et al.* Systematic exploration of synergistic drug pairs. *Mol Syst Biol* 2011;7:544.

Diaconis, P.a.G., Ronald L. Spearman's footrule as a measure of disarray. *Journal of the Royal Statistical Society. Series B (Methodological)* 1977:262-268.

Geyer, C.E., *et al.* Lapatinib plus capecitabine for HER2-positive advanced breast cancer. *N Engl J Med* 2006;355(26):2733-2743.

Goswami, C.P., *et al.* A New Drug Combinatory Effect Prediction Algorithm on the Cancer Cell Based on Gene Expression and Dose-Response Curve. *CPT Pharmacometrics Syst Pharmacol* 2015;4(2):e9.

Guimera, R. and Sales-Pardo, M. A Network Inference Method for Large-Scale Unsupervised Identification of Novel Drug-Drug Interactions. *Plos Comput Biol* 2013;9(12):e1003374.

Held, M.A., *et al.* Genotype-Selective Combination Therapies for Melanoma Identified by High-Throughput Drug Screening. *Cancer Discov* 2013;3(1):52-67.

Huang, J.L., *et al.* Systematic Prediction of Pharmacodynamic Drug-Drug Interactions through Protein-Protein-Interaction Network. *Plos Comput Biol* 2013;9(3).

Huang, L., *et al.* DrugComboRanker: drug combination discovery based on target network analysis. *Bioinformatics* 2014;30(12):228-236.

Iorio, F., *et al.* Discovery of drug mode of action and drug repositioning from transcriptional responses. *Proc Natl Acad Sci U S A* 2010;107(33):14621-14626.

Iorio, F., Tagliaferri, R. and di Bernardo, D. Identifying network of drug mode of action by gene expression profiling. *J Comput Biol* 2009;16(2):241-251.

Irizarry, R.A., *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 2003;4(2):249-264.

Jia, J., *et al.* Mechanisms of drug combinations: interaction and network perspectives. *Nat Rev Drug Discov* 2009;8(2):111-128.

Joensuu, H., *et al.* Effect of the tyrosine kinase inhibitor STI571 in a patient with a metastatic gastrointestinal stromal tumor. *N Engl J Med* 2001;344(14):1052-1056.

Kim, S., *et al.* PubChem Substance and Compound databases. *Nucleic Acids Res* 2016;44(D1):D1202-1213.

Li, P., *et al.* Large-scale exploration and analysis of drug combinations. *Bioinformatics* 2015;31(12):2007-2016.

Lin, S. Space oriented rank-based data integration. *Stat Appl Genet Mol Biol* 2010;9:Article20.

LoRusso, P.M., *et al.* Accelerating Cancer Therapy Development: The Importance of Combination Strategies and Collaboration. Summary of an Institute of Medicine Workshop. *Clin Cancer Res* 2012;18(22):6101-6109.

Maggiora, G., *et al.* Molecular Similarity in Medicinal Chemistry. *J Med Chem* 2014;57(8):3186-3204.

Pang, K.F., *et al.* Combinatorial therapy discovery using mixed integer linear programming. *Bioinformatics* 2014;30(10):1456-1463.

Risso, D., *et al.* Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat Biotechnol* 2014;32(9):896-902.

Subramanian, A., *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;102(43):15545-15550.

Vilar, S., *et al.* Drug-drug interaction through molecular structure similarity analysis. *J Am Med Inform Assn* 2012;19(6):1066-1074.

Vogel, C.L., *et al.* Efficacy and safety of trastuzumab as a single agent in first-line treatment of HER2-overexpressing metastatic breast cancer. *J Clin Oncol* 2002;20(3):719-726.

Yang, J., *et al.* DIGRE: Drug-Induced Genomic Residual Effect Model for Successful Prediction of Multidrug Effects. *CPT Pharmacometrics Syst Pharmacol* 2015;4(2):e1.

Zhao, X.M., *et al.* Prediction of Drug Combinations by Integrating Molecular and Pharmacological Data. *Plos Comput Biol* 2011;7(12).