OXFORD

Briefings in Bioinformatics, 2025, **26(1)**, bbaf054
https://doi.org/10.1093/bib/bbaf054
Problem Solving Protocol

iDOMO: identification of drug combinations via multi-set operations for treating diseases

Xianxiao Zhou[1,2,3,*], Ling Wu[4], Minghui Wang[1,2,3], Guojun Wu[4], Bin Zhang[1,2,3,5,*]

[1]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York, NY 10029, United States
[2]Mount Sinai Center for Transformative Disease Modeling, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York, NY 10029, United States
[3]Icahn Genomics Institute, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York, NY 10029, United States
[4]Barbara Ann Karmanos Cancer Institute, Department of Oncology, Wayne State University School of Medicine, 4100 John R, Detroit, MI 48201, United States
[5]Department of Pharmacological Sciences, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York, NY 10029, United States

*Corresponding author. Xianxiao Zhou. E-mail: xianxiao.zhou@mssm.edu; Bin Zhang. E-mail: binzhang.ucla@gmail.com

Abstract

Combination therapy has become increasingly important for treating complex diseases which often involve multiple pathways and targets. However, experimental screening of drug combinations is costly and time-consuming. The availability of large-scale transcriptomic datasets (e.g. CMap and LINCS) from *in vitro* drug treatment experiments makes it possible to computationally predict drug combinations with synergistic effects. Towards this end, we developed a computational approach, termed Identification of Drug Combinations via Multi-Set Operations (iDOMO), to predict drug synergy based on multi-set operations of drug and disease gene signatures. iDOMO quantifies the synergistic effect of a pair of drugs by taking into account the combination's beneficial and detrimental effects on treating a disease. We evaluated iDOMO, in a DREAM Challenge dataset with the matched, pre- and post-treatment gene expression data and cell viability information. We further evaluated the performance of iDOMO by concordance index and Spearman correlation on predicting the Highest Single Agency (HSA) synergy scores for four most common cancer types in two large-scale drug combination databases, showing that iDOMO significantly outperformed two existing popular drug combination approaches including the Therapeutic Score and the SynergySeq Orthogonality Score. Application of iDOMO to triple-negative breast cancer (TNBC) identified drug pairs with potential synergistic effects, with the combination of trifluridine and monobenzone being the most synergistic. Our in vitro experiments confirmed that the top predicted drug combination exerted a significant synergistic effect in inhibiting TNBC cell growth. In summary, iDOMO is an effective method for the *in silico* screening of synergistic drug combinations and will be a valuable tool for the development of novel therapeutics for complex diseases.

Keywords: drug repurposing; drug combination; multiple set operations; synergistic effect; drug gene signature; disease gene signature

Introduction

Genetic and phenotypic heterogeneity among cancer patients causes substantial drug resistance [1–3]. Therapies with combinations of two or more drugs are designed to overcome drug resistance, under the assumption that cancer cells resistant to one drug might be responsive to another one. A study analysed many effective and approved combinations concluding that therapeutic benefit is likely to arise from independent drug action rather than drug additivity or synergy [4]. Currently, available data from clinical studies suggest the best treatment outcome is achieved by a combination of targeted drugs with conventional chemotherapy [5, 6]. However, the identification of successful drug cocktails is not a simple task and is hampered by the lack of standards in terminology, experimental protocols and models as well as data analysis. Given millions of possible drug combinations from thousands of drugs, it is crucial to accurately predict effective drug combinations using various functional genomic data *in silico*. This is necessary for the identification and successful implementation of combination therapy. Generally, both similar and distinct gene expression signatures can lead to the synergy of drug combinations [7]. Several approaches have been developed to predict drug synergy from gene expression-derived drug signatures [8–13], protein–protein interaction network-based [14, 15] or deep-learning/machine-learning [15–19]. However, those approaches did not consider disease signatures, a synergistic combination of drug signatures does not necessarily mean it is beneficial for treating a specific disease. For example, Therapeutic Score (TS) [12] was developed to predict synergy score based on disease signature between disease and control, and drug signatures of before and after treatment. However, TS doesn't take the drug signature size into consideration, while drugs inducing a large number of non-treatment-related gene expression changes imply significant side effects. Another study developed SynergySeq Orthogonality Score (OS) [20] from transcriptional consensus signatures (TCS) across multiple replicates and cell lines. However, the TCS was obtained from various model systems, which cannot exactly measure drug responses in each specific cell line. Two drugs are likely to have preferred synergistic effect if the combination of their signatures have enhanced beneficial effect and reduced side effect.

Transcriptomic data of drug perturbations such as the Connectivity Map (CMap) [21] and the Library of Integrated

Received: July 31, 2024. **Revised:** December 10, 2024. **Accepted:** January 27, 2025

Network-based Cellular Signatures (LINCS) project [22] have been widely used to search for potential novel indications through matching signatures of diseases and drugs/compounds based on genome-wide gene expression phenotypes. Under the assumption that drugs that can reverse gene expression alterations in diseases would have potential therapeutic effect on restoring disease states to healthy conditions, we previously developed a drug repositioning algorithm EMUDRA [23] based on drug-induced gene signatures and disease gene signatures. On top of individual drugs predicted by methods like EMUDRA, this study develops a new approach, termed 'Identification of Drug Combinations via Multi-Set Operations (iDOMO)', to prioritize drug combinations using multi-set operations based quantification of a combination's potential synergistic effect using gene expression signatures of diseases and drugs. We systematically compared iDOMO with two existing drug combination methods, TS and SynergySeq OS, using two drug combination databases including DrugCombDB and DrugComb. The two databases contain drug combination tests from high-throughput screening studies, external databases and manual curations and in particular, they include experimentally tested data from over 1 million drug combinations across >100 cell lines/tissues.

Results

Development of the iDOMO approach

We developed a new approach iDOMO to identify drug combinations by modeling each drug pair's beneficial and detrimental effects on diseased patients based on drug and disease gene expression signatures. Specifically, the up and down-regulated genes in a disease are determined by comparing gene expression data in a disease-relevant tissue from patients with the disease and those from healthy controls while the up and down-regulated genes for a drug are identified by comparing gene expression data in model systems from treatment experiments with the drug and those from control experiments. Next, these six up or down-regulated gene signatures from the disease and two drugs are further divided into 27 classes by multi-set operations including intersection and difference (Fig. 1A). Specifically, drug or drug-combination-induced gene expression changes that are opposite to the changes in the disease are defined as beneficial changes, while drug- or drug-combination-induced changes that are consistent with those in the disease are detrimental (Fig. 1B). In addition, gene expression changes that are induced by two drugs but in opposite directions are neutral. Then, the significance of the overlap among 3 gene sets of each class could be calculated using the Super Exact Test (SET) [24] $P(x \geq |A \cap B \cap C|) = 1 - \sum_{x=0}^{x-1} \mathbf{Pr}(|A \cap B| = j; |A| = a, |B| = b)\ \mathbf{Pr}(|A \cap B \cap C| = x; |A \cap B| = j, |C| = c)$ (Fig. 1B). Scores of beneficial classes and detrimental classes (see Methods for details) are summed up as beneficial scores (*BS*) and harmful scores (HS), respectively. Similarly, for each drug in a combination, an individual drug score (IDS) can be calculated for each drug based on Fisher's exact test (FET) to evaluate its overall beneficial effect. Lastly, the final score for the drug combination, termed total drug combination score (TDCS), is calculated as the difference between the drug combination scores (*BS* and *HS*);

$$TDCS\left(d, g_A, g_B\right) = BS - HS$$

and a combination synergy score (CSS), is calculated as the difference between the TDCS and the maximum of the IDS against a disease signature:

$$CSS\left(d, g_A, g_B\right) = BS - HS - \mathbf{max}\left(IDS_A, IDS_B\right).$$

Performance evaluation through simulations

We further designed a procedure to evaluate the robustness of TDCS. First, we generated disease signatures by 1) combining up-regulated genes in a target drug pair as disease-down-regulated genes, and 2) combining down-regulated genes in the same drug pair as disease-up-regulated genes. Then, we added random genes as noise to the disease signature and then calculated TDCS for each drug combination. If the drug pair is ranked at the top with various noise levels, it means the approach can robustly identify the right drug combination for a disease signature. We performed 1000 simulations for each noise rate level. As shown in Fig. 2, the average rank is less than 10 with 35% or less of random genes in a signature. However, the average rank is significantly increased with more than 70% or more random genes added. The result indicates that the drug combination approach can robustly identify drug combinations.

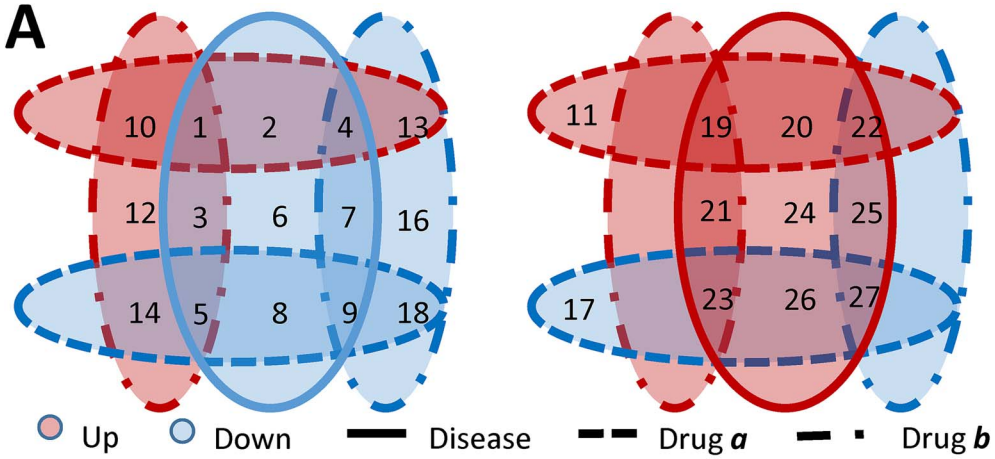
Performance evaluation of iDOMO and comparison with two existing approaches

We then evaluated the performance of iDOMO based on a DREAM challenge dataset in which cell viability was measured for 91 drug pairs of 14 individual drugs in a human B-cell lymphoma (BCL) cell line OCI-LY3 [9]. Gene expression data were also generated from the samples before and after treatment of each drug. We first identified the signatures of the forteen drugs from the gene expression data by comparing the samples after treatment to those before treatment. We further identified a BCL signature from a B-cell lymphoma gene expression dataset (GSE72026) [25]. The BCL signature is composed of 3816 up-regulated genes and 2378 down-regulated genes from the comparison of the BCL samples and the peripheral blood B cell control. Finally, we applied our SET approach to the BCL signature and the forteen drug signatures to calculate synergy scores for the 91 drug pairs.

As in the DREAM challenge study [9], we compared our prediction score of the 91 drug pairs and their matched EOB, which measures the drug pair's inhibition effect that excess additive effect in experiments (see Methods). A drug pair with EOB = 0 has an additive effect, whereas a drug pair with positive (or negative) EOB values has a synergistic (or antagonistic) effect. As shown in Fig. 3, TDCS has a significant positive correlation with the observed EOB (Spearman correlation $\rho = 0.40$, *P*-value = 8.32E-05). Furthermore, we also assessed the performance of iDOMO using a probabilistic concordance index (PC-index), which was used as the main performance measurement in the DREAM challenge study [9]. The iDOMO has a PC-index of 0.6207, which is better than the best approach, the Drug Induced Genomic Residual Effect (DIGRE) model (PC-index = 0.6130) which was tested in the DREAM challenge paper [9].

We further evaluated the performance of the iDOMO in the publicly available drug combination screening databases, DrugCombDB [26] and DrugComb [27]. We first extracted the Highest Single Agency (HSA) model-based drug combination synergy score, for four widely used cell lines (MCF7, PC3, HT29, and A549), corresponding to the top four cancer types with the highest number of new cases in the US [38230766], i.e. breast carcinoma (BRCA), prostate adenocarcinoma (PRAD), colorectal adenocarcinoma (COAD) and lung adenocarcinoma (LUAD). For each of the four cell lines, we then matched the drug

A



B

ID	d	$-g_a$	$-g_a * d$	$-g_b$	$-g_b * d$	$(-g_a * d) * (-g_b * d)$	P
1	-1	-1	***B***	-1	***B***	***BB***	p_1
2	-1	-1	***B***	0	***N***	***B***	p_2
3	-1	0	***B***	-1	***N***	***B***	p_3
4	-1	-1	***B***	1	***H***	***N***	NA
5	-1	1	***B***	-1	***H***	***N***	NA
6	-1	0	***N***	0	***N***	***N***	NA
7	-1	0	***N***	1	***H***	***H***	p_7
8	-1	1	***N***	0	***H***	***H***	p_8
9	-1	1	***H***	1	***H***	***HH***	p_9
10	0	-1	***H***	-1	***H***	***HH***	p_{10}
11	0	-1	***H***	0	***N***	***H***	p_{11}
12	0	0	***H***	-1	***N***	***H***	p_{12}
13	0	-1	***H***	1	***H***	***N***	NA
14	0	1	***H***	-1	***H***	***N***	NA
15	0	0	***N***	0	***N***	***N***	NA
16	0	0	***N***	1	***H***	***H***	p_{16}
17	0	1	***N***	0	***H***	***H***	p_{17}
18	0	1	***H***	1	***H***	***HH***	p_{18}
19	1	-1	***H***	-1	***H***	***HH***	p_{19}
20	1	-1	***H***	0	***N***	***H***	p_{20}
21	1	0	***H***	-1	***N***	***H***	p_{21}
22	1	-1	***H***	1	***B***	***N***	NA
23	1	1	***H***	-1	***B***	***N***	NA
24	1	0	***N***	0	***N***	***N***	NA
25	1	0	***N***	1	***B***	***B***	p_{25}
26	1	1	***N***	0	***B***	***B***	p_{26}
27	1	1	***B***	1	***B***	***BB***	p_{27}

Figure 1. Evaluation of the effect of the combination of two drugs on a disease of interest. (A) Venn diagram of 27 subsets from the multi-set operations of the disease and individual drug signatures. Left panel shows the Venn diagram of down-regulated genes in a disease signature and two drug signatures of a combination. The right panel shows the Venn diagram of up-regulated genes in a disease signature and two drug signatures in combination. Red indicates up-regulated genes in a signature and blue indicates down-regulated genes. (B) Scoring 27 subsets of the disease signature d and opposite drug signatures $-g_A$ and $-g_B$. -1 indicates the gene subsets down-regulated in the disease signature or down-regulated in the drug signatures, while 1 indicates the gene subsets up-regulated in the disease signature or down-regulated in the drug signatures, and 0 indicates no significant change subsets in signatures. $-g_A * d$ indicates the drug effect of drug a on the disease d, and $-g_B * d$ indicates the drug effect of drug b on the disease d, while $(-g_A * d) * (-g_B * d)$ indicates the effects of drugs a and b on disease d. P indicates the p-value calculated by the super exact test. *B* indicates a beneficial overlap between the disease and drug signatures, while *N* indicates neutral changes and *H* indicates a detrimental (harmful) overlap between disease and drug signatures.

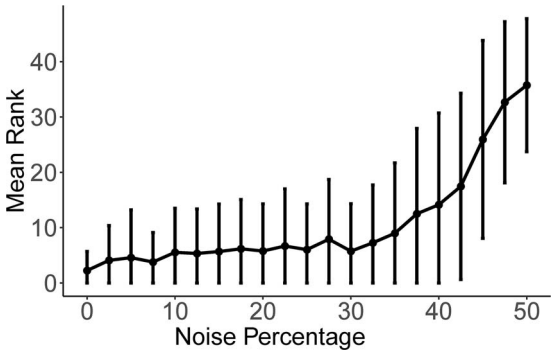


Figure 2. Robustness of the proposed drug combination method iDOMO. Two drugs were randomly chosen as a target true combination. Disease signatures were generated by sampling genes from the signatures of the target drug pair and then adding a certain number of random genes that are not in the drug signatures. The mean ranks and standard deviations were based on 1000 repeats.

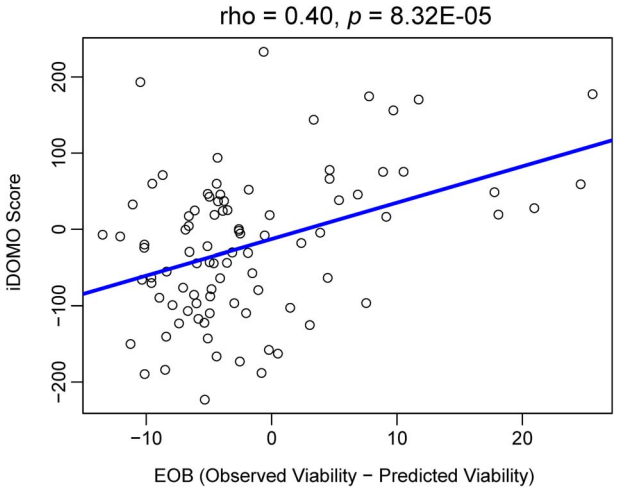


Figure 3. Performance evaluation of the iDOMO in the DREAM challenge dataset. The observed excess over bliss (EOB) data was downloaded from the DREAM challenge paper [9]. iDOMO scores were calculated based on the 91 pairs of the 14 individual drug signatures and the B-cell lymphoma signature identified from the GSE72026 [25].

combinations to the drugs tested in the LINCS and extracted the combinations with both drugs tested in the LINCS. Next, to calculate the iDOMO drug combination scores for those drug combinations matched to LINCS, we identified the drug signatures between the drug-treated samples and vehicle control samples for the four cell lines. Furthermore, we identified cancer signatures for the corresponding four cancer types using the gene expression samples from the TCGA under FDR < 0.05 and fold change >2, which were summarized in Supplementary Table S1, which is available online at http://bib.oxfordjournals.org/. Finally, we calculated the iDOMO score using the cancer signature and drug signatures for the matched drug combinations for each of the four cancer types. We then used the concordance index (C-index) and Spearman correlation to assess the performance of the iDOMO scores on the prediction of the HSA synergy score. As shown in Table 1, the C-index is about 0.60 across the four cancer types in both the DrugCombDB and DrugComb databases. Furthermore, the iDOMO TDCS is significantly positively correlated with the HSA synergy score (Fig. 4). For instance, the C-index of the BRCA is 0.6093 for the prediction of the 550 drug combinations in the DrugCombDB, and the iDOMO TDCS is significantly correlated with the HSA synergy score (Spearman's $\rho = 0.3243$, $P = 6.20E\text{-}15$) (Fig. 4A). Moreover, iDOMO performed better for BRCA in the DrugComb. We matched 597 drug combinations for the LINCS MCF7 in the DrugComb database. The C-index of the iDOMO TDCS is 0.6431 and Spearman's $\rho = 0.4200$ ($P < 2.20E\text{-}16$) (Fig. 4B). We further tested iDOMO TDCS with the cancer signatures identified by FDR < 0.05 and fold change >1.5. The performance (Supplementary Table S2 available online at http://bib.oxfordjournals.org/ and Fig. S2) is similar to the cancer signatures based on FDR < 0.05 and fold change >2, suggesting the iDOMO TDCS is robust to the thresholds for signature identification. Furthermore, we tested the CSS in the same databases using the same signatures. The performance is slightly inferior to the TDCS but still good (Supplementary Table S3, which is available online at http://bib.oxfordjournals.org/).

We also compared the performance of our iDOMO TDCS and two existing gene-signature-based approaches TS score [12] and SynergySeq Orthogonality Score (TCS OS) [20], in predicting the drug combination synergy score HSA in the DrugComb and DrugCombDB databases. We first matched the drug combination TS and TCS OS values of the four TCGA cancer types (BRCA, PRAD, COAD, and LUAD) to the HSA synergy scores of the corresponding cancer type cell lines (MCF7, PC3, HT29, and A549) of the same drug combinations in the two databases. We then calculated each cancer type's C-index and Spearman correlation coefficient between TS or TCS OS prediction and HSA synergy score. Generally, TS and TCS OS showed worse performance than our iDOMO TDCS measured by the C-index of predicting the HSA synergy score and the correlation between the prediction scores and the corresponding HSA observations (Supplementary Table S4 available online at http://bib.oxfordjournals.org/, Fig. 4C and Fig. S3), except for TS prediction for the PRAD in the DrugComb database. Among the three approaches, TCS OS is the worst. For example, the TS values extracted from the BRCA signatures were not correlated with the HSA scores in the DrugCombDB when the same drug combinations were considered (Spearman's $\rho = 0.0876$, $P = 0.691$, Fig. S3A and Table S4). The C-index was 0.5079 as a random prediction (Table S4). Notably, the C-indices of TCS OS were very close to 0.5 (Fig. 4C and Table S5) and Spearman's correlation coefficients were close to 0 (Fig. S4) except for the prediction for COAD (Table S5). Therefore, the extensive performance comparison demonstrates iDOMO's superior performance over two popular approaches for predicting drug combinations.

In summary, iDOMO outperforms the existing approaches for predicting drug combinations in both the DREAM challenge and public drug combination databases.

Application of iDOMO to triple negative breast cancer

We then applied iDOMO to predict drug combinations for treating triple negative breast cancer (TNBC). In our previous EMUDRA study [23], we predicted 20 drugs for TNBC using a TNBC signature of 4776 differentially expressed genes (DEGs) between the TNBC samples and the matched controls in the TCGA cohort [28]. We calculated the TDCS for the 190 combinations of the 20 drugs using the 4776 DEGs and drug signatures identified from the CMap [21]. As shown in Table 2, the combination of two compounds, trifluridine and monobenzone, was predicted to be most effective with a SET combination score of 123.63. Notably, trifluridine is a component in 7 of the predicted top 10 most synergistic drug pairs, while monobenzone is a component in 3 of the 10 drug pairs. As shown in Fig. 5, trifluridine and monobenzone significantly reversed the TNBC signature individually and in combination.

Table 1. The performance of iDOMO total drug combination score on the prediction of experimental HSA synergy score.

Database	Signature*	Cell line	# Drug Combs	C-Index	Rho	P-value
DrugCombDB	BRCA	MCF7	550	0.6093	0.3243	6.20E-15
	PRAD	PC3	614	0.5578	0.1655	3.79E-05
	COAD	HT29	299	0.5897	0.2641	3.66E-06
	LUAD	A549	246	0.6052	0.3139	5.03E-07
DrugComb	BRCA	MCF7	597	0.6431	0.4200	< 2.20E-16
	PRAD	PC3	652	0.5681	0.2016	2.10E-07
	COAD	HT29	367	0.6187	0.3443	1.18E-11
	LUAD	A549	248	0.6056	0.3113	5.69E-07

[a]Cancer signatures were identified based on adjusted P-value <0.05 and folder change >2.

Table 2. The top 10 pairs of compounds that are predicted to be synergistic for treating TNBC.

Drug A	Drug B	BS-HS	IDS_A	IDS_B	DCS
Trifluridine	Monobenzone	123.63	40.94	26.74	55.95
Trifluridine	0173570–0000	113.16	40.94	21.1	51.12
Monobenzone	0173570–0000	95.97	26.74	21.1	48.13
Trifluridine	Ciclopirox	107.34	40.94	24.25	42.15
Trifluridine	Vidarabine	106.54	40.94	24.84	40.76
Trifluridine	Methotrexate	115.68	40.94	35.23	39.51
Resveratrol	Trifluridine	101.76	30.42	40.94	30.4
Etoposide	Monobenzone	110.79	64.53	26.74	19.52
Etoposide	Vidarabine	99.41	64.53	24.84	10.04
Etoposide	Trifluridine	110.98	64.53	40.94	5.51

Experimental validation of the top predicted drug combination

As the combination of trifluridine and monobenzone was predicted by iDOMO to be most effective for treating TNBC, we further experimentally validated the drug pair *in vitro*. We first calculated the correlation coefficients of the gene expression levels of the untreated TNBC cell lines in CCLE [29] and the primary TNBC samples from TCGA [28]. Among the cell lines that showed significant correlation with most of the TCGA primary TNBC patient samples (Fig. S5), SUM159 is the only cell line available in our lab. So, we chose SUM159 for our *in vitro* validation experiment cell to validate the synergistic effects of trifluridine and monobenzone. Trifluridine is a fluorinated thymidine analog, which can incorporate into DNA and inhibit DNA replication leading to the death of cancer cells [30]. The exact mechanism of action of monobenzone is not fully understood, but it may inhibit cancer cell growth by inhibiting ribonucleotide reductase [31] and inducing oxidative stress and inflammatory cytokines [32]. Thus, the two drugs may have synergistic anti-tumor effects by targeting different pathways.

We treated SUM159 cells with trifluridine or monobenzone or a combination (ratio 1:2) at seven concentrations. Compared to the dimethyl sulfoxide (DMSO) control, the cell number was significantly decreased with trifluridine or monobenzone treatment at each of the seven concentrations (Fig. 6A and B). Moreover, the combination of the 2 drugs showed a synergistic effect at the lowest four doses (T25 + M50, T12.5 + M25, T6.25 + M12.5, T3.125 + M6.25) (Fig. 6C and D). We quantitatively analyzed the synergism of the combination of trifluridine and monobenzone using Chou-Talalay's combination index method. The results indicated that the combination exerted a synergistic effect when cell growth inhibition was less than 50%, with the combination index (CI) ranging from 0.785 to 0.577 for the fraction affected (Fa) ranging from 0.694 to 0.896, as shown in the Fa-CI plot (Fig. 6C). Thus, this experiment validated our prediction that there is a synergistic effect in the combination of trifluridine and monobenzone.

Discussion

In this study, we developed a new method iDOMO to predict the synergistic effect of drug combinations based on disease and drug gene expression signatures. Based on the concept that reversing disease gene expression changes would reverse disease-related phenotypes, we modeled the beneficial and detrimental effects of a combination using the drug and disease gene signatures. Several tools have been developed to predict drug synergy using machine learning methods [33–35]. However, those tools require multiple experimental data points to calculate dose–response matrices. It is costly and time-consuming to generate dose–response matrices as it requires drug treatment experiments across multiple dosages and time points. iDOMO takes into account all possible combinations of the up- and down-regulated gene signatures of disease and two drugs for computing the final combination score. Thus, the synergy of drug combinations can be predicted for disease signatures derived from massive drug treatment data, e.g. the CMap and LINCS. In the CMap, drug combination can be calculated among over 1300 drugs in three cell lines, while in the LINCS dataset, our approach can be applied to over 12 000 FDA-approved and investigational compounds in over 70 cell lines. Nevertheless, gene expression change patterns are essential for drug synergy but not sufficient to detect a synergy in a specific disease [7]. In theory, it is ideal to target different parts of a disease network or pathway with multiple drugs to maximally normalize the dysregulated network or pathway [14]. On the other hand, concurrent inhibition of parallel subnetworks or pathways that drive proliferation and anti-apoptosis can acquire the best efficiency for cancer treatment [36, 37].

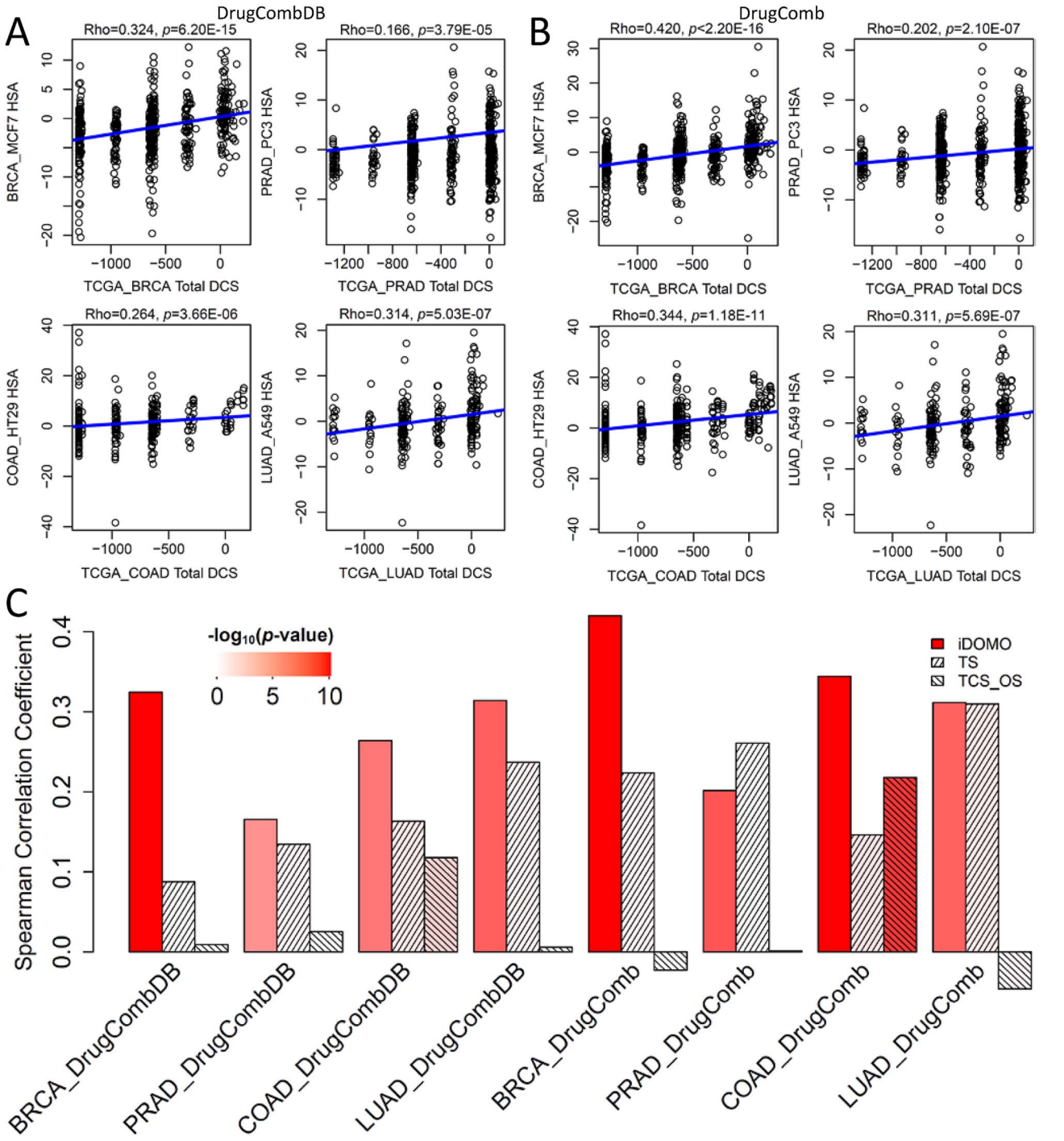
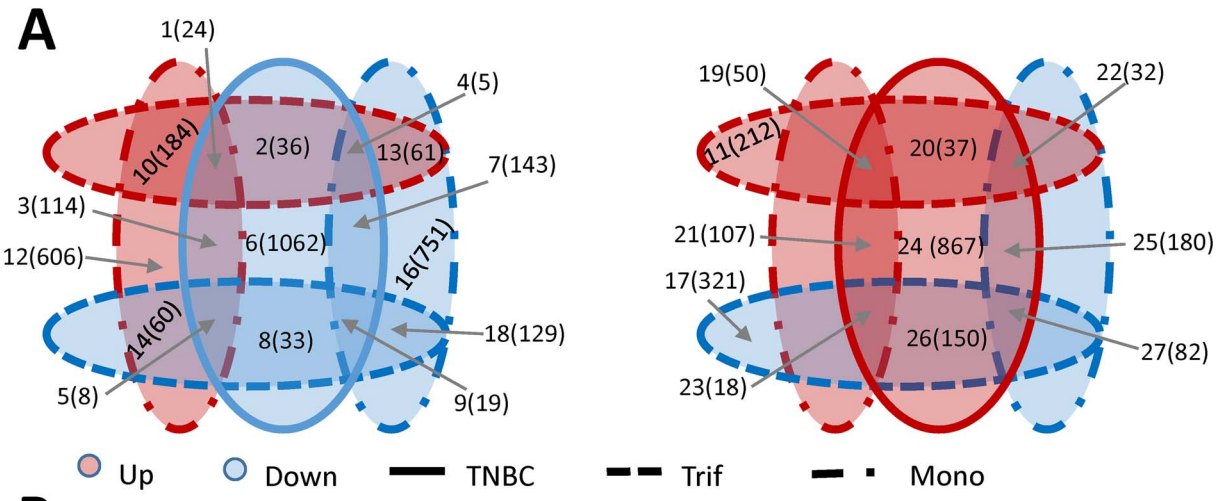


Figure 4. Performance evaluation of iDOMO in the DrugCombDB and DrugComb databases and in comparison with two existing approaches, therapeutic score (TS) and TCS based SynergySeq orthogonality score (TCS OS). Scatter plots of the total drug combination score (TDCS) and highest single agency (HSA) synergy score for four TCGA cancer types (BRCA, PRAD, COAD and LUAD) in the DrugCombDB (A) and DrugComb (B) databases. (C) Comparison of the correlation coefficients between HSA synergy scores and prediction scores of iDOMO TDCS, TS and TCS OS.

Application of iDOMO to TNBC identified the combination of trifluridine and monobenzone as a potential novel therapy for treating TNBC. Trifluridine is a thymidine-based nucleoside analog, which is incorporated into DNA via phosphorylation to inhibit DNA replication and cell proliferation. Trifluridine has been used as an antiviral agent to treat herpetic keratitis [38]. Currently, it is usually co-administrated with tipiracil, a thymidine phosphorylase inhibitor, to treat metastatic colorectal and gastric cancers [36, 39]. Monobenzone is a potent skin-bleaching agent used as a treatment for over-pigmentation in the clinic [40]. It was also reported that monobenzone is used to treat cutaneous metastatic melanoma patients in a phase 2 trial [37]. Recently, monobenzone was shown to be a potent inhibitor for lysine-specific demethylase 1 (KDM1A) [41], which is an oncogene that promotes the initiation and development of various cancers via diverse cellular signaling pathways [42, 43]. Monobenzone can inhibit the proliferation and migration of various cancer cells by competitively inhibiting KDM1A reversibly [41]. In our analysis,



B

ID	d	$-Trif$	$Trif_d$	$-Mono$	$Mono_d$	$Combn$	$\#Genes$	GO enrichment
1	-1	-1	*B*	-1	*B*	*BB*	24	memory
2	-1	-1	*B*	0	*N*	*B*	36	G1/S transition of mitotic cell cycle
3	-1	0	*B*	-1	*N*	*B*	114	response to cAMP
4	-1	-1	*B*	1	*H*	*N*	5	membranous septum morphogenesis
5	-1	1	*B*	-1	*H*	*N*	8	hemopoiesis
6	-1	0	*N*	0	*N*	*N*	1062	angiogenesis
7	-1	0	*N*	1	*H*	*H*	143	renal system development
8	-1	1	*N*	0	*H*	*H*	33	cellular localization
9	-1	1	*H*	1	*H*	*HH*	19	regulation of biosynthetic process
10	0	-1	*H*	-1	*H*	*HH*	184	response to endoplasmic reticulum stress
11	0	-1	*H*	0	*N*	*H*	212	protein folding
12	0	0	*H*	-1	*N*	*H*	606	cellular protein modification process
13	0	-1	*H*	1	*H*	*N*	61	protein folding
14	0	1	*H*	-1	*H*	*N*	60	apoptotic signaling pathway
15	0	0	*N*	0	*N*	*N*	NA	NA
16	0	0	*N*	1	*H*	*H*	751	mRNA splice site selection
17	0	1	*N*	0	*H*	*H*	321	RNA processing
18	0	1	*H*	1	*H*	*HH*	129	lactation
19	1	-1	*H*	-1	*H*	*HH*	50	response to endoplasmic reticulum stress
20	1	-1	*H*	0	*N*	*H*	37	regulation of catalytic activity
21	1	0	*H*	-1	*N*	*H*	107	tRNA aminoacylation for protein translation
22	1	-1	*H*	1	*B*	*N*	32	mitotic cell cycle process
23	1	1	*H*	-1	*B*	*N*	18	cellular nitrogen compound metabolic process
24	1	0	*N*	0	*N*	*N*	867	DNA metabolic process
25	1	0	*N*	1	*B*	*B*	180	cell cycle
26	1	1	*N*	0	*B*	*B*	150	ncRNA metabolic process
27	1	1	*B*	1	*B*	*BB*	82	DNA replication and Cell cycle

Figure 5. SET analysis of the drug pair trifluridine (Trif) and monobenzone (mono). (A) Venn diagram of 27 subsets from the multi-set operations of the triple-negative breast cancer (TNBC) signature and Trif and mono signatures. The left panel shows the Venn diagram of down-regulated genes in the TNBC signature and the up- and down-regulated genes of the Trif and mono signatures. The right panel shows the Venn diagram of up-regulated genes in the TNBC signature and the up- and down-regulated genes of the Trif and mono signatures. Red indicates up-regulated genes in a signature and blue indicates down-regulated genes. (B) Scoring 27 subsets of the TNBC signature d and drug signatures trifluridine and monobenzone. The GO enrichment column shows the most significant enriched GO biological processes of the genes shared by the TNBC and drug signatures.

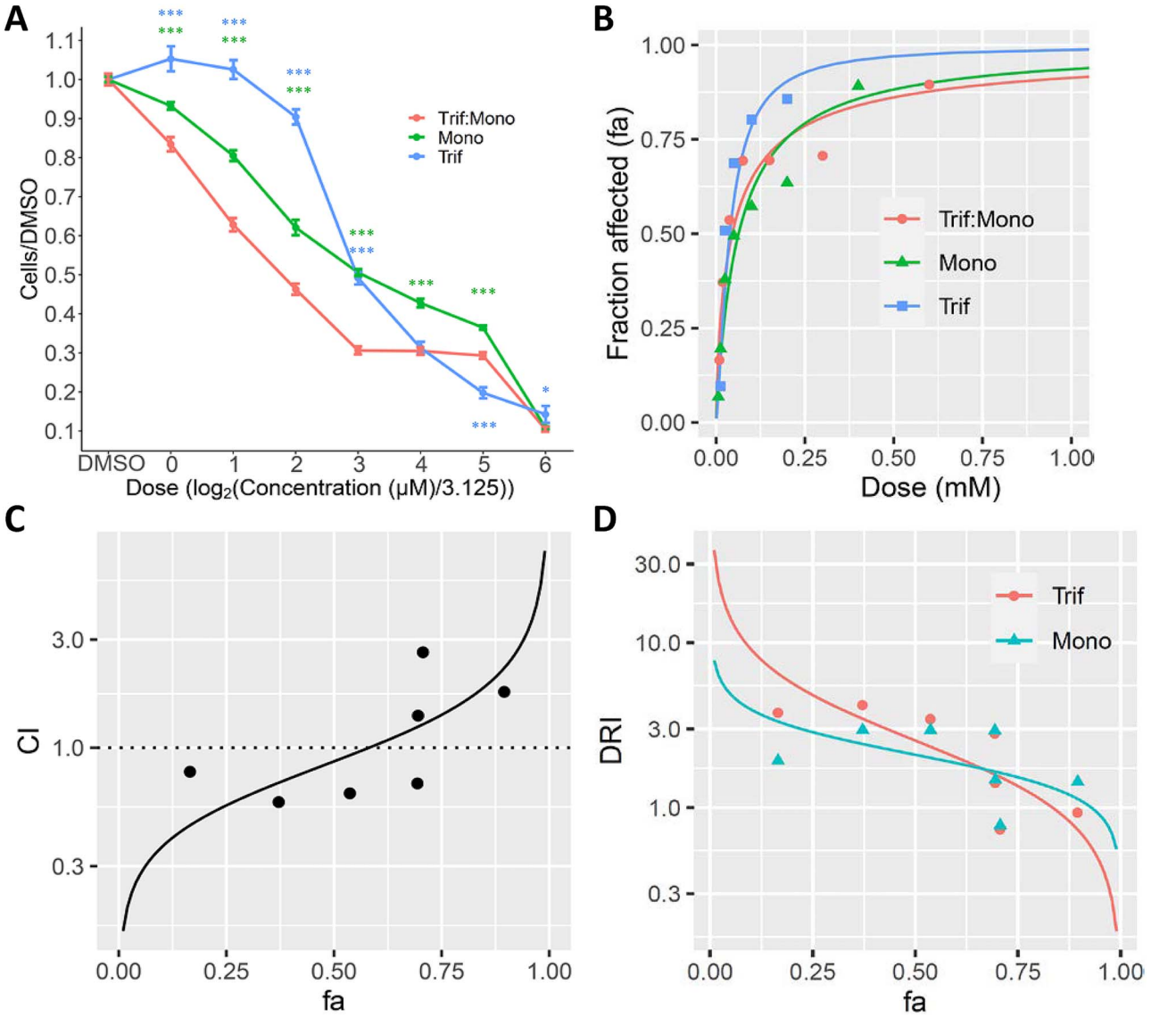


Figure 6. Validation of the top predicted drug combination for treating TNBC. (A) the cell effect of trifluridine (multiple of 3.125 μM), monobenzone (multiple of 6.25 μM) and their combination (ratio 1:2) at 7 concentrations compared to DMSO baseline. The differences between individual drugs and combination were tested by t-test and shown by markers *** ($P < 0.0001$) and * ($P < 0.01$). (B) the dose-effect plot of trifluridine, monobenzone and their combination. The effect is shown by the fraction affected (FA). (C) the FA-CI (combination index) plot of the trifluridine and monobenzone combination at a 1:2 ratio. CI < 1 indicates synergy; CI = 1 indicates additive; CI >1 indicates antagonism. CI is shown in the log10 scale. (D) Dose-reduction index (DRI) (see methods) plot with different FAs, DRI is shown in the log10 scale.

both trifluridine and monobenzone can down-regulate the gene expression of DNA replication, DNA repair and mitotic cell cycle. Furthermore, trifluridine can down-regulate the gene expression of ribonucleoprotein complex biogenesis, tRNA processing and rRNA processing, while monobenzone down-regulated more DNA replication and cell cycle-related genes as well as microtubule-based process, RNA methylation and protein ubiquitination genes. Besides, monobenzone up-regulated genes of response to cAMP and cytokine biosynthesis, which may increase the immune response to cancer cells. The above functional analysis of the genes induced by the trifluridine and monobenzone combination suggested that the synergy may be achieved by targeting the same pathways from different directions and different pathways essential to cancer. Additional pathways may be targeted by the combination of the two drugs. Nevertheless, drug synergy-induced gene expression changes should be measured to explore the synergistic mechanisms of the combination.

Previous studies have suggested that synergy is condition-specific [44–46]. Cell type, disease subtype, mutation and gene expression status may influence the synergy of drug combinations. In the drug combination DREAM Challenge, the results from the best-performed algorithms suggested that the similarity between the gene expression changes from monotherapies is associated with synergy [9]. A recent study showed that synergistic combinations could generate unique transcriptomic changes that were not induced by any of the monotherapies [7]. This study also demonstrated that the correlation of transcriptomic response of monotherapies was necessary but not sufficient for drug synergy [7]. In this study, we have shown that comparison of drug-gene signatures and different parts of a disease signature can be used to predict synergy with improved performance (PC-index) than the algorithms developed through the DREAM Challenge. Several factors may affect the performance of iDOMO. First, it is essential to identify the gene

signature of a disease under study from the primary tissue(s) responsible for the disease. Second, selecting disease-relevant cell lines in LINCS is crucial, as different cell lines may exhibit varied molecular signatures in response to drug treatment. This selection can be achieved by comparing the expression profiles of LINCS cell lines with those from disease-causing tissues, allowing us to choose the most similar cell lines for drug repurposing. Third, identifying optimal drug combinations from a large pool of drugs is computationally intensive. Thus, optimizing iDOMO to quickly screen drug combinations is necessary. Lastly, the current performance evaluation is based only on a limited set of drug combinations available in both LINCS and the DrugCombDB or DrugComb databases. The drug repurposing research community should collaborate to generate larger datasets, including matched pre- and post-treatment expression data and viability information across many cell lines for future studies. With the pre- and post-treatment data from the CMap and LINCS phase I and phase II, our approach could be an efficient and cost-effective method for *in silico* screening of synergistic drug combinations for a variety of diseases such as cancers and neurodegenerative diseases.

Methods

Data preprocessing and normalization

The microarray data of drug and DMSO treatments were downloaded from the CMap [21]. Raw data were preprocessed as described in the EMUDRA paper [23]. Briefly, raw intensities of probe sets were preprocessed using the robust multi-array average (RMA) procedure [47] and summarized to a gene-level expression matrix of 12 328 genes. The expression matrix was then corrected for batch effects using a linear model [48]. For the LINCS dataset, the normalized level 3 data matrices of compound treated and control samples were downloaded from the GSE92742 (LINCS v1.1) of the Gene Expression Omnibus (GEO) database [49] and the LINCS data portal (LINCS beta) (https://clue.io/releases/data-dashboard). The LINCS data matrix includes the expression levels of directly measured landmark genes and the gene expression data inferred from those directly measured genes. In total, 12 328 genes were included in the analysis of this study. The data matrices were also corrected for batch effects. The level 3 RNA-seq data of TNBC samples and adjacent normal samples were downloaded from The Cancer Genome Atlas (TCGA) data portal [50].

The RNA-Seq gene expression data for the four cancer types (BRCA, PRAD, COAD, and LUAD) were downloaded from TCGA.

Signature identification

To identify a signature for each drug/compound of a cell line in the CMap, the gene expression profiles of experiments treated by various doses and treatment durations were combined and compared to DMSO control profiles of the same batch. *P*-values of differential expression were calculated using a linear model from the R package *limma* [37]. *Limma* uses an Empirical Bayes moderated t-statistic to calculate *p*-values with modified standard error and degrees of freedom. The standard error is moderated across genes with borrowed information from all genes to improve the inference of any single gene. The degrees of freedom are also adjusted using a factor representing a priori degrees of freedom for the model. To correct the multiple-testing problem, *P*-values were adjusted using the Benjamini–Hochberg (BH) procedure [51]. Drug signatures were then identified using the thresholds of adjusted *P*-value <0.05. Analogously, drug signatures of the LINCS dataset were also identified using the same procedure and thresholds for each cell line. As each cell line was treated with each drug in six different concentrations in more than a dozen batches, those profiles were treated as replicates and put together to identify one high-quality drug-induced signature for each cell line. Similarly, the TNBC signature was also identified using the *limma* package [48] by comparing the TNBC samples to the matched normal controls in TCGA. The signatures of the four cancer types were also identified by comparing the cancer and normal samples using the linear model in the *limma* package [48] and *P*-values were corrected with the BH procedure [51]. Significant differentially expressed genes (DEGs) of the signatures were filtered by FDR < 0.05 and FC > 2 or 1.5.

Drug combination scores based on detrimental and beneficial effects

We propose a SET-based drug combination score (DCS) to model the overall effect of a disease signature and drug signatures. This DCS evaluates how the different intersections of the drug and disease signatures exert a beneficial or detrimental effect on the disease under study. Specifically, the DCS is based on the following assumptions: (i) a change (up- or down-regulation) induced by one or both drugs which are opposite to that induced by a disease under study is beneficial; (ii) an expression change in a combination is detrimental if the change is in the same direction as that induced by the disease; (iii) a change that is in the opposite directions by two drugs is neutral (Fig. 1B).

Suppose there are three gene sets A, B, and C with a, b and c genes, respectively. Suppose j genes overlap between A and B, and x genes overlap between A, B, and C. The probability of the intersection is exact x can be calculated using the following equation:

$$\begin{aligned}\mathbf{Pr}\left(|A\cap B\cap C|=\ x;|A|=a,|B|=b,|C|=c\right)\\ =\ \mathbf{Pr}\left(|A\cap B|=j;|A|=a,|B|=b\right)\\ \times\mathbf{Pr}\left(|A\cap B\cap C|=x;|A\cap B|=j,|C|=c\right)\end{aligned}$$

Note that

$$0\le j\le \mathbf{min}\left(a-x,b-x,c\right) \quad (1)$$

Suppose that there is a disease signature d comprised of the up-regulated gene set d_{up}, down-regulated gene set d_{dn} and gene set d_0 without significant change. Similarly, a drug signature g_A (g_B) includes the up-regulated gene set $g_{A_{up}}$ ($g_{B_{up}}$), the down-regulated gene set $g_{A_{dn}}$ ($g_{B_{dn}}$), and the unchanged gene set g_{A_0} (g_{B_0}). As shown in Fig. 1, there are 27 subsets from the multi-set operations on the aforementioned gene signatures. Given three gene sets, the probability of obtaining no less than a given number of overlapping genes can be calculated using the Equation 1 [24]. The *P*-values of the six subsets with beneficial effects can be calculated as follows:

$$p_1=\mathbf{Pr}\left(x\ge|d_{dn}\cap g_{A_{up}}\cap g_{B_{up}}|\right)$$

$$p_2=\mathbf{Pr}\left(x\ge|d_{dn}\cap g_{A_{up}}\cap g_{B_0}|\right)$$

$$p_3=\mathbf{Pr}\left(x\ge|d_{dn}\cap g_{A_0}\cap g_{B_{up}}|\right)$$

$$p_{25}=\mathbf{Pr}\left(x\ge|d_{up}\cap g_{A_0}\cap g_{B_{dn}}|\right)$$

$$p_{26}=\mathbf{Pr}\left(x\ge|d_{up}\cap g_{A_{dn}}\cap g_{B_0}|\right)$$

$$p_{27}=\mathbf{Pr}\left(x\ge|d_{up}\cap g_{A_{dn}}\cap g_{B_{dn}}|\right)$$

The p_1, p_2, p_3, p_{25}, p_{26} and p_{27} can be summarized into a beneficial score (BS) in Equation 2 as follows:

$$BS = \sum_{k=1}^{3} -\log_{10}(p_k) + \sum_{k=25}^{27} -\log_{10}(p_k) \quad (2)$$

Similarly, the gene subsets with detrimental effects can be calculated using equation 1 as follows:

$$p_7 = \Pr\left(x \geq |d_{dn} \cap g_{A_0} \cap g_{B_{dn}}|\right)$$

$$p_8 = \Pr\left(x \geq |d_{dn} \cap g_{A_{dn}} \cap g_{B_0}|\right)$$

$$p_9 = \Pr\left(x \geq |d_{dn} \cap g_{A_{dn}} \cap g_{B_{dn}}|\right)$$

$$p_{10} = \Pr\left(x \geq |d_0 \cap g_{A_{up}} \cap g_{B_{up}}|\right)$$

$$p_{11} = \Pr\left(x \geq |d_0 \cap g_{A_{up}} \cap g_{B_0}|\right)$$

$$p_{12} = \Pr\left(x \geq |d_0 \cap g_{A_0} \cap g_{B_{up}}|\right)$$

$$p_{16} = \Pr\left(x \geq |d_0 \cap g_{A_0} \cap g_{B_{dn}}|\right)$$

$$p_{17} = \Pr\left(x \geq |d_0 \cap g_{A_{dn}} \cap g_{B_0}|\right)$$

$$p_{18} = \Pr\left(x \geq |d_0 \cap g_{A_{dn}} \cap g_{B_{dn}}|\right)$$

$$p_{19} = \Pr\left(x \geq |d_{up} \cap g_{A_{up}} \cap g_{B_{up}}|\right)$$

$$p_{20} = \Pr\left(x \geq |d_{up} \cap g_{A_{up}} \cap g_{B_0}|\right)$$

$$p_{21} = \Pr\left(x \geq |d_{up} \cap g_{A_0} \cap g_{B_{up}}|\right)$$

Then detrimental score (HS) can be summarized in Equation 3 by the sum of p_7, p_8, p_9, p_{10}, p_{11}, p_{12}, p_{16}, p_{17}, p_{18}, p_{19}, p_{20} and p_{21} as follows:

$$HS = \sum_{k=7}^{12} -\log_{10}(p_k) + \sum_{k=16}^{21} -\log_{10}(p_k) \quad (3)$$

4, 5, 6, 13, 14, 15, 22, 23 and 24 are neutral (neither beneficial nor detrimental). Therefore, the *BS* and *HS* can be summarized into aTDCS in Equation 4:

$$TDCS(d, g_A, g_B) = BS - HS \quad (4)$$

Similarly, we defined a score for each drug signature based on the *P*-values the FET of the overlap between the drug signature and a disease signature. As shown in Supplemental Fig. S1, the significance of the overlap between individual drug signature and disease signature was log 10 transformed and added up in Equation 5 as follows to obtain IDS against a disease signature.

$$IDS_i = -\log_{10}(p_{i1}) - \log_{10}(p_{i3}) + \log_{10}(p_{i2}) + \log_{10}(p_{i4}) + \log_{10}(p_{i6}) + \log_{10}(p_{i8}) \quad (5)$$

Finally, the *BS*, *HS* and IDSs are summarized into a TDCS and a combination synergy score (CSS) in Equation 6:

$$CSS(d, g_A, g_B) = BS - HS - \max(IDS_A, IDS_B) \quad (6)$$

A combination is predicted to be synergistic if the TDCS and CSS are larger than 0.

Combinatorial therapeutic score and TCS based OS

Combinatorial therapeutic score (TS) scores of BRCA, PRAD, COAD, and LUAD were extracted from the Supplementary Table S10 of the TS paper [12]. Drug combinations with both drug names that matched the drug combinations in the databases DrugComb and DrugCombDB were further extracted to evaluate the performance of the TS in the 4 cancer types for the prediction of HSA.

Transcriptional Consensus Signature (TCS) based OS was calculated for the TCGA BRCA, PRAD, COAD and LUAD signatures using the OS algorithm [20] and TCS from the LINCS L1000 data, as described in the original SynergySeq study [20]. Specifically, TCS were calculated from the LINCS L1000 dataset, including the small molecule treatment transcriptional data of the MCF7, PC3, HA1E, HCC515, VCAP, A375, HEPG2, HT29, A549, ASC, NEU, and NPC cell lines. Aggregation was performed for each cell line first. By measuring gene expression data before and after drug perturbation, a gene was included in the aggregated expression profile of a cell line if it had a mean |z-score| > 1 across the samples using both the same small molecule and the same cell line. Next, the aggregated gene expression profiles were collapsed for each the small molecule across all cell lines to produce the TCS. A gene was included in the final TCS of a small molecule across cell lines if it was consistently up/down-regulated by a |z-score| > 1 in more than 30% of the cell lines that were treated by the same small molecule. The concordance ratio (CR) and the disease discordance ratio (DR) were then calculated for each drug pair [20]. Herein, CR is defined as the ratio of TCS genes of a compound/small molecule that in the same direction as the reference TCS gene signature to the ones in the opposite directions (Equation 7). DR is defined as the ratio of TCS genes induced by a drug in an opposite direction to the ones in the same direction against a disease signature, giving that the disease signature genes that are missing in the reference signature (Equation 8):

$$\mathrm{CR} = \frac{\sum_{i=1}^{\#genes}[a_i]}{\sum_{i=1}^{\#genes}[b_i]} \text{ with } a_i = \begin{cases} 1, & if\ z_i \cdot r_i > 0 \\ 0, & if\ z_i \cdot r_i < 0 \end{cases} \text{ and } b_i = \begin{cases} 1, if\ z_i \cdot r_i < 0 \\ 0, if\ z_i \cdot r_i > 0 \end{cases} \quad (7)$$

$$\mathrm{DR} = \frac{\sum_{i=1}^{\#genes}[b_i \cdot c_i]}{\sum_{i=1}^{\#genes}[a_i \cdot c_i]} \text{ with } a_i = \begin{cases} 1, & if\ z_i \cdot d_i > 0 \\ 0, & if\ z_i \cdot d_i < 0 \end{cases}, \quad b_i = \begin{cases} 1, & if\ z_i \cdot d_i < 0 \\ 0, & if\ z_i \cdot d_i > 0 \end{cases} \text{ and } c_i = \begin{cases} 1, & if\ r_i = 0 \\ 0, & if\ r_i \neq 0 \end{cases} \quad (8)$$

where *z*, *d*, and *r* denote the TCS signatures of a small molecule/compound, the disease and a reference compound/small molecule, respectively. After scaling CR with DR into the range [0 and 1], the orthogonality score (OS) of each compound to a reference compound is defined as (Equation 9).

$$\mathrm{OS} = \sqrt{(1 - \mathrm{CR})^2 + \mathrm{DR}^2} \quad (9)$$

Drug combination data of DrugCombDB and DrugComb databases

The drug combination data with experimentally tested synergy scores were downloaded from the DrugComb and DrugCombDB databases. Combinations with HSA synergy scores were extracted. For the TS drug combination prediction in the Table S10 of the

TS paper [12], HSA synergy scores of corresponding combinations with both drugs matched to the combinations in each of the two databases were extracted to evaluate the performance of the TS. For the iDOMO TDCS, CSS and SynergySeq OS scores, HSA synergy scores of drug combinations with both drugs mapped to compound/small molecule in the LINCS data were extracted to evaluate their performance based on prediction scores.

Gene ontology enrichment analysis for drug-induced genes

Gene ontology (GO) analysis [52] was performed for each gene subset in Fig. 1 to obtain the top biological processes that were significantly enriched with the subset genes. The analysis was based on the GO.db and org.Hs.eg.db packages of the Bioconductor [53]. *P*-values of enrichment were calculated by the hypergeometric distribution using an *R* package GO-function [54] and corrected by the Benjamini–Yekutieli procedure [55]. Redundant biological processes were removed through the procedures in the GO-function [54].

Performance evaluation by simulation with drug signatures identified from the CMap

Target drug pairs were randomly generated from the top 20 drugs predicted for the TNBA signature using the EMUDRA [23]. For each pair, a fake disease signature was generated by replacing the genes in the drug signatures induced by the golden standard drug pair with a proportion of noise genes. Those noise genes were generated from the complementary gene set of the drug signatures that were not changed by the two drugs. Then, the fake disease signature was used to test iDOMO and obtain the rank for all the drug pairs. The rank of the golden standard drug pair was used to assess the performance. This procedure was repeated 1000 times under each noise rate (proportion). The average rank and standard deviation are reported in Fig. 2.

Testing in the DREAM challenge dataset

The signatures of the 14 drugs in the DREAM Challenge dataset [9] were identified using the same procedure for identifying the CMap signatures as described above. Briefly, the gene expression data of 282 samples of the drug and DMSO-treated B-Cell Lymphoma cell line OCI-LY3 were downloaded from the GEO accession GSE51068 [9]. The probes were normalized using the RMA procedure using the Bioconductor package in R to obtain log2 transformed probeset expression levels. Probesets were then annotated to gene symbols using the platform Affymetrix Human Genome U219 Array (HG-U219). Probesets annotated to none or multiple gene symbols were discarded. Probesets annotated to the same gene symbol were averaged to obtain gene-level expression data. Lowly expressed genes with the mean expression levels at the bottom 20% were removed from further analysis. In total, 16 747 genes were included in the analysis. For each drug, differential expression analysis was conducted by comparing profiles of various concentrations and treatment duration time of the same drug to the DMSO-treated profiles. *P*-values were calculated using the Empirical Bayes moderated t-statistic of the R package *limma* [48]. BH procedure [51] was applied to the *P*-values to correct the multiple testing problem.

The B-cell lymphoma signature was identified from the gene expression profiles of the GEO dataset GSE72026 [25]. Normalized and log2-transformed gene expression data matrix was downloaded from the GEO database. Probesets were annotated to gene symbols using the platform Agilent-039494 SurePrint G3 Human GE v2 8x60K Microarray 039381 (GPL17077). Probesets annotated to lincRNA were removed from further analysis. As described above, probeset-level expression values were summarized to gene-level expression values using the same procedure as the HG-U219. In total, 23 675 genes were included in the following analysis. Differential expression analysis was conducted using the R package *limma* [48] and *P*-values were corrected by the BH procedure [51]. Under a 5% FDR, 4671 up-regulated and 3544 down-regulated genes were identified by comparing 4 intravascular large B-cell lymphoma samples to four peripheral blood B cell control samples. After removing the genes that are not on the HG-U219 platform, 3816 up-regulated and 2378 down-regulated genes were used in the drug synergy prediction.

We used the probabilistic concordance index (PC-index) to evaluate the performance of the prediction, which is modified from the concordance index (C-index). C-index computes the proportion of concordance between the predicted SET scores and observed Excess over Bliss (EOB) ranks of drug pairs to quantify the performance of the prediction. If drugs D_x and D_y have an additive effect, the expected fractional inhibition f_{add} of their combination should be:

$$f_{add} = 1 - (1 - f_x) \times (1 - f_y)$$

where f_x and f_y are experimentally measured fractional inhibitions of drug D_x and D_y, respectively. EOB is determined by subtracting the expected fractional inhibition f_{add} from the fractional inhibition induced by the drug combination (of D_x and D_y) f_{xy}.

$$EOB = f_{xy} - f_{add}$$

Therefore, a drug pair with 0 EOB has an additive effect, while a drug pair with positive (or negative) EOB values has a synergistic (or antagonistic) effect. The experimentally measured EOB data of the 91 drug pairs and their standard error of the mean (S.E.M.) were downloaded from the DREAM Challenge dataset [9].

To calculate the C-index, the 91 drug pairs of the 14 drugs were first ranked by the experimentally determined EOB, from the most synergistic to the additive and then the most antagonistic. Suppose r_i and r_j are the observed ranks of drug pairs i and j, and pr_i and pr_j are the predicted ranks of the same drug pairs i and j. The concordance score C_{ij} of the drug pairs i and j is 1 if the observed ranks and predicted ranks are concordant; otherwise $C_{ij} = 0$.

$$C_{ij} = \begin{cases} 1, & if \ (r_i > r_j \ and \ pr_i > pr_j \ or \ r_i < r_j \ and \ pr_i < pr_j) \\ 0, & if \ (r_i > r_j \ and \ pr_i < pr_j \ or \ r_i < r_j \ and \ pr_i > pr_j) \end{cases}$$

The C-index is then defined as follows:

$$C.index = \frac{2}{90 \times 91} \sum_{i=1}^{90} \sum_{j=i+1}^{91} C_{ij}$$

Considering the deviations of the drug inhibition experiments, the c-index is modified to the PC-index as in the DREAM Challenge paper [9].

Validation experiments in TNBC cell line SUM159

SUM159 cells were cultured in F12-HAM's medium with 5% FBS and supplemented with 5 μg/ml insulin and 1 μg/ml hydrocortisone. SUM159 cells were seeded into 96 wells at a density of 4x10^3/well. The next day, different doses of trifluridine (200,

100, 50, 25, 12.5, 6.25, and 3.125 μM) and monobenzone (400, 200, 100, 50, 25, 12.5, and 6.25 μM) were added either individually or in combination (ratio 1:2). The cells were then treated for 72 hr before being fixed with ice-cold trichloroacetic acid for 1 hr at 4°C. The cells were then washed with water and air-dried. About 100 μl of 0.057% sulforhodamine B (w/v) was added to stain the cells for 30 min at RT. The cells were then rinsed quickly with 1% acetic acid for four times to remove the unbound dye. The plates were air-dried. About 200 μl of 10 mM Tris base solution (pH 10.5) was added to each well and shaken for 30 min at RT to ensure the dye was homogeneously dissolved.

Then OD510 was recorded with a plate reader [56]. The synergistic effect of the two drugs was calculated using Compusyn following the Chou-Talalay method [57]. The synergistic effect was shown as CI values. (CI < 1 synergy, CI = 1 additive, CI > 1 antagonism). The CompuSyn analysis involved determining the CI and dose-reduction index (DRI) values for each drug combination tested at a consistent dose ratio. CI values reflect the combined effect of multiple drugs (synergistic, additive, or antagonistic). In contrast, DRI values signify fold decrease in drug dose required in a combination to achieve the same efficacy (Fa) as the individual drug. Under this context, drug combinations with CI values < 1 are deemed synergistic, and DRI values ≥ 1 are considered favorable, mainly due to concerns about toxicity associated with combining multiple drugs. To evaluate drug combinations for cancer treatment, these criteria are especially crucial if met at high effect (Fa) levels, where the goal is to eliminate a significant proportion of tumor cells. Hence, we designated any drug combination with CI < 1 and DRI ≥ 1 (for both drugs in the combination) at Fa ranging from 0.694 to 0.896 as a promising combination.

Key Points

- Previous drug combination synergy prediction approaches did not take into account differences across diseases.
- A new computational approach iDOMO was developed to predict synergistic effects of drug combinations based on multi-set operations of drug and disease gene signatures.
- iDOMO significantly outperforms two popular methods for predicting drug combinations.
- iDOMO was applied to triple-negative breast cancer and the top predicted drug combination was experimentally validated.

Acknowledgements

This work was supported in part by the computational resources and staff expertise provided by Scientific Computing at the Icahn School of Medicine at Mount Sinai.

Supplementary data

Supplementary data is available at *Briefings in Bioinformatics* online.

Conflict of interest: None declared.

Funding

This work was supported in part by grants from the National Institutes of Health (NIH)/National Institute on Aging [R01AG085182, RF1AG074010, U01AG046170, RF1AG057440, R01AG057907, U01AG052411, R01AG062355, U01AG058635, RF1AG054014, RO1AG068030, and R56AG058655], and NIH/National Institute of Diabetes and Digestive and Kidney Diseases [R01DK118243].

Data availability

The data underlying this article are available in the GitHub, at https://github.com/zxx59834/iDOMO

References

1. Turajlic S, Sottoriva A, Graham T. *et al.* Resolving genetic heterogeneity in cancer. *Nat Rev Genet* 2019;**20**:404–16.
2. Chatterjee N, Bivona TG. Polytherapy and targeted cancer drug resistance. *Trends Cancer* 2019;**5**:170–82.
3. Vlachostergios PJ, Faltas BM. Treatment resistance in urothelial carcinoma: an evolutionary perspective. *Nat Rev Clin Oncol* 2018;**15**:495–509.
4. Palmer AC, Sorger PK. Combination cancer therapy can confer benefit via patient-to-patient variability without drug additivity or synergy. *Cell* 2017;**171**:1678–1691 e1613.
5. Cunningham D, Humblet Y, Siena S. *et al.* Cetuximab monotherapy and cetuximab plus irinotecan in irinotecan-refractory metastatic colorectal cancer. *N Engl J Med* 2004;**351**:337–45.
6. Geyer CE, Forster J, Lindquist D. *et al.* Lapatinib plus capecitabine for HER2-positive advanced breast cancer. *N Engl J Med* 2006;**355**:2733–43.
7. Diaz JE, Ahsen ME, Schaffter T. *et al.* The transcriptomic response of cells to a drug combination is more than the sum of the responses to the monotherapies. *Elife* 2020;**9**:e52707.
8. Dar RD, Hosmane NN, Arkin MR. *et al.* Screening for noise in gene expression identifies drug synergies. *Science* 2014;**344**:1392–6.
9. Bansal M, Yang J, Karan C. *et al.* A community computational challenge to predict the activity of pairs of compounds. *Nat Biotechnol* 2014;**32**:1213–22.
10. Li X, Xu Y, Cui H. *et al.* Prediction of synergistic anti-cancer drug combinations based on drug target network and drug induced gene expression profiles. *Artif Intell Med* 2017;**83**:35–43.
11. Matlock K, Berlow N, Keller C. *et al.* Combination therapy design for maximizing sensitivity and minimizing toxicity. *BMC Bioinformatics* 2017;**18**:116.
12. Huang CT, Hsieh CH, Chung YH. *et al.* Perturbational gene-expression signatures for combinatorial drug discovery. *iScience* 2019;**15**:291–306.
13. Kong W, Midena G, Chen Y. *et al.* Systematic review of computational methods for drug combination prediction. *Comput Struct Biotechnol J* 2022;**20**:2807–14.
14. Cheng F, Kovacs IA, Barabasi AL. Network-based prediction of drug combinations. *Nat Commun* 2019;**10**:1197.
15. Yang J, Xu Z, Wu WKK. *et al.* GraphSynergy: a network-inspired deep learning model for anticancer drug combination prediction. *J Am Med Inform Assoc* 2021;**28**:2336–45.
16. Li J, Tong XY, Zhu LD. *et al.* A machine learning method for drug combination prediction. *Front Genet* 2020;**11**:1000.
17. Ianevski A, Giri AK, Gautam P. *et al.* Prediction of drug combination effects with a minimal set of experiments. *Nat Mach Intell* 2019;**1**:568–77.
18. Zhang T, Zhang L, Payne PRO. *et al.* Synergistic drug combination prediction by integrating multiomics data in deep learning models. *Methods Mol Biol* 2021;**2194**:223–38.

19. Wu L, Wen Y, Leng D. *et al.* Machine learning methods, databases and tools for drug combination prediction. *Brief Bioinform* 2022;**23**:1–21.
20. Stathias V, Jermakowicz AM, Maloof ME. *et al.* Drug and disease signature integration identifies synergistic combinations in glioblastoma. *Nat Commun* 2018;**9**:5315.
21. Lamb J, Crawford ED, Peck D. *et al.* The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 2006;**313**:1929–35.
22. Subramanian A, Narayan R, Corsello SM. *et al.* A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* 2017;**171**:1437–1452 e1417.
23. Zhou X, Wang M, Katsyv I. *et al.* EMUDRA: Ensemble of Multiple Drug Repositioning Approaches to improve prediction accuracy. *Bioinformatics* 2018;**34**:3151–9.
24. Wang M, Zhao Y, Zhang B. Efficient test and visualization of multi-set intersections. *Sci Rep* 2015;**5**:16923.
25. Shimada K, Shimada S, Sugimoto K. *et al.* Development and analysis of patient-derived xenograft mouse models in intravascular large B-cell lymphoma. *Leukemia* 2016;**30**:1568–79.
26. Liu H, Zhang W, Zou B. *et al.* DrugCombDB: a comprehensive database of drug combinations toward the discovery of combinatorial therapy. *Nucleic Acids Res* 2020;**48**:D871–81.
27. Zheng S, Aldahdooh J, Shadbahr T. *et al.* DrugComb update: a more comprehensive drug sensitivity data repository and analysis portal. *Nucleic Acids Res* 2021;**49**:W174–84.
28. Cancer genome atlas N: comprehensive molecular portraits of human breast tumours. *Nature* 2012;**490**:61–70.
29. Barretina J, Caponigro G, Stransky N. *et al.* The cancer cell line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 2012;**483**:603–7.
30. Matsuoka K, Iimori M, Niimi S. *et al.* Trifluridine induces p53-dependent sustained G2 phase arrest with its massive misincorporation into DNA and few DNA strand breaks. *Mol Cancer Ther* 2015;**14**:1004–13.
31. Dong J, Zhong T, Xu Z. *et al.* Identification of monobenzone as a novel potential anti-acute myeloid leukaemia agent that inhibits RNR and suppresses tumour growth in mouse xenograft model. *Cancers (Basel)* 2022;**14**:4710, 1–18.
32. Khalid-Meften A, Liaghat M, Yazdanpour M. *et al.* The effect of monobenzone cream on oxidative stress and its relationship with serum levels of IL-1beta and IL-18 in vitiligo patients. *J Cosmet Dermatol* 2024;**23**:4085–93.
33. Fan K, Cheng L, Li L. Artificial intelligence and machine learning methods in predicting anti-cancer drug combination effects. *Brief Bioinform* 2021;**22**:bbab271, 1–12.
34. Preuer K, Lewis RPI, Hochreiter S. *et al.* DeepSynergy: predicting anti-cancer drug synergy with deep learning. *Bioinformatics* 2018;**34**:1538–46.
35. Torkamannia A, Omidi Y. Ferdousi R: a review of machine learning approaches for drug synergy prediction in cancer. *Brief Bioinform* 2022;**23**:bbac075, 1–19.
36. Kang C, Dhillon S, Deeks ED. Trifluridine/Tipiracil: a review in metastatic gastric cancer. *Drugs* 2019;**79**:1583–90.
37. Teulings HE, Tjin EPM, Willemsen KJ. *et al.* Anti-melanoma immunity and local regression of cutaneous metastases in melanoma patients treated with monobenzone and imiquimod; a phase 2 a trial. *Onco Targets Ther* 2018;**7**:e1419113, 1–14.
38. Carmine AA, Brogden RN, Heel RC. *et al.* Trifluridine: a review of its antiviral activity and therapeutic use in the topical treatment of viral eye infections. *Drugs* 1982;**23**:329–53.
39. Burness CB, Duggan ST. Trifluridine/Tipiracil: A review in metastatic colorectal cancer. *Drugs* 2016;**76**:1393–402.
40. van den Boorn JG, Melief CJ, Luiten RM. Monobenzone-induced depigmentation: from enzymatic blockade to autoimmunity. *Pigment Cell Melanoma Res* 2011;**24**:673–9.
41. Ma P, Jia G, Song Z. Monobenzone, a novel and potent KDM1A inhibitor, suppresses migration of gastric cancer cells. *Front Pharmacol* 2021;**12**:640949, 1–7.
42. Chen Y, Ren B, Yang J. *et al.* The role of histone methylation in the development of digestive cancers: a potential direction for cancer management. *Signal Transduct Target Ther* 2020;**5**:143.
43. Jia G, Cang S, Ma P. *et al.* Capsaicin: a "hot" KDM1A/LSD1 inhibitor from peppers. *Bioorg Chem* 2020;**103**:104161, 1–5.
44. Lehar J, Krueger AS, Avery W. *et al.* Synergistic drug combinations tend to improve therapeutically relevant selectivity. *Nat Biotechnol* 2009;**27**:659–66.
45. Holbeck SL, Camalier R, Crowell JA. *et al.* The National Cancer Institute ALMANAC: a comprehensive screening resource for the detection of anticancer drug pairs with enhanced therapeutic activity. *Cancer Res* 2017;**77**:3564–76.
46. Held MA, Langdon CG, Platt JT. *et al.* Genotype-selective combination therapies for melanoma identified by high-throughput drug screening. *Cancer Discov* 2013;**3**:52–67.
47. Irizarry RA, Hobbs B, Collin F. *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 2003;**4**:249–64.
48. Ritchie ME, Phipson B, Wu D. *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015;**43**:e47, 1–13.
49. Barrett T, Wilhite SE, Ledoux P., *et al*: NCBI GEO: archive for functional genomics data sets–update. *Nucleic Acids Res* 2013, **41**:D991–5.
50. The Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* 2012;**490**:61–70.
51. Benjamini Y, Hochberg Y. Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J Roy Stat Soc B Met* 1995;**57**:289–300.
52. Ashburner M, Ball CA, Blake JA. *et al.* Gene ontology: tool for the unification of biology. *The Gene Ontology Consortium Nat Genet* 2000;**25**:25–9.
53. Gentleman RC, Carey VJ, Bates DM. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 2004;**5**:R80.
54. Wang J, Zhou X, Zhu J. *et al.* GO-function: deriving biologically relevant functions from statistically significant functions. *Brief Bioinform* 2012;**13**:216–27.
55. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. *Ann Stat* 2001;**29**:1165–88.
56. Vichai V, Kirtikara K. Sulforhodamine B colorimetric assay for cytotoxicity screening. *Nat Protoc* 2006;**1**:1112–6.
57. Chou TC, Talalay P. Quantitative analysis of dose-effect relationships: the combined effects of multiple drugs or enzyme inhibitors. *Adv Enzyme Regul* 1984;**22**:27–55.

Briefings in Bioinformatics, 2025, **26(1)**, bbaf054
https://doi.org/10.1093/bib/bbaf054
Problem Solving Protocol