SynergyFinder Plus: Toward Better Interpretation and Annotation of Drug Combination Screening Datasets

Shuyu Zheng, Wenyu Wang, Jehad Aldahdooh, Alina Malyutina, Tolou Shadbahr, Ziaurrehman Tanoli, Alberto Pessia, Jing Tang*

Research Program in Systems Oncology, Faculty of Medicine, University of Helsinki, Helsinki 00290, Finland

Running title: *Zheng S et al / SynergyFinderPlus for Drug Combination Data Analysis*

* Corresponding author.
E-mail: jing.tang@helsinki.fi (Tang J).

Word count: 3742
Reference count: 50
Figure count: 6
Supplementary figure count: 3
Supplementary table count: 1
Count of letters in the article title: 91
Count of letters in the running title: 47
Count of keywords: 5
Count of words in Abstract: 204

Abstract

Combinatorial therapies have been recently proposed to improve the efficacy of anticancer treatment. The SynergyFinder R package is a software used to analyze pre-clinical drug combination datasets. Here, we report the major updates to the SynergyFinder R package for improved interpretation and annotation of drug combination screening results. Unlike the existing implementations, the updated SynergyFinder R package includes five main innovations. (1) We extend the mathematical models to higher-order drug combination data analysis and implement dimension reduction techniques for visualizing the synergy landscape. (2) We provide a statistical analysis of drug combination synergy and sensitivity with confidence intervals and *P* values. (3) We incorporate a synergy barometer to harmonize multiple synergy scoring methods to provide a consensus metric for synergy. (4) We evaluate drug combination synergy and sensitivity to provide an unbiased interpretation of the clinical potential. (5) We enable fast annotation of drugs and cell lines, including their chemical and target information. These annotations will improve the interpretation of the mechanisms of action of drug combinations. To facilitate the use of the R package within the drug discovery community, we also provide a web server at www.synergyfinderplus.org as a user-friendly interface to enable a more flexible and versatile analysis of drug combination data.

Keywords:

SynergyFinder; Drug combinations; Synergy modeling; Drug discovery; Drug combination sensitivity analysis

Introduction

Many complex diseases including cancer and infectious diseases develop drug resistance [1,2]. To achieve more sustainable clinical efficacy, multi-targeted drug combinations have been proposed to tackle disease signaling pathways more systematically [3]. With the advances in high-throughput drug screening technologies, an increasing number of drug combinations are being tested in multiple disease models, such as cancer cell lines and patient-derived primary cultures. These preclinical drug combination experiments are conducted to identify the most synergistic and effective drug combinations that result in improved cellular responses, such as cell viability and toxicity. With the potential hits prioritized from screening, more functional studies are warranted to elucidate the mechanisms of action of the drug interactions, leading to the identification of drug combination response biomarkers that are critical for stratifying patients for more targeted therapies.

For the evaluation of the potential of a drug combination, mathematical and statistical models are needed to characterize the expectation if the drugs are not interactive; thereafter the boosting effects of the drug combination can be formally quantified. There are currently four major synergy models: highest single agent (HSA) [4], Loewe additivity (LOEWE) [5], Bliss independence (BLISS) [6], and zero interaction potency (ZIP) [7]. However, the overall effects of a drug combination may not lead to sufficient efficacy, despite a strong degree of interaction. Therefore, it is recommended to evaluate both synergy (*i.e.,* the degree of interactions) and sensitivity (*i.e.,* the overall treatment efficacy) simultaneously [8]. Previously developed tools that can analyze drug combination synergy include COMPUSYSN [9], BRAID [10], Combenfit [11], SynergyFinder [12], and Synergy [13]. However, these tools are designed mainly for analyzing two-drug combinations only. More recently, the SynergyFinder2 tool [14] has been developed, which extends to three-drug combinations. However, only the HSA model was appropriately constructed for its implementation. The key differences in the modeling are summarized in Table S1. Furthermore, the visualization of high-order drug combinations becomes non-trivial, as none of the existing visualization methods can deal with combinations of three or more drugs [15]. More importantly, there is a lack of implementation tools that can harmonize multiple synergy models to derive a consensus about the degree of interactions [16].

To address these limitations, we provide a major update of the SynergyFinder R package that enables novel functions, including (1) formal mathematical models to assess the synergy of high-order drug combinations and visualization of the degree of synergy using a dimension reduction

technique; (2) formal statistical methods to evaluate the significance of synergy; (3) implementation of the synergy barometer to systematically compare the results from different synergy models; and (4) implementation of a synergy-sensitivity (SS) plot to evaluate the potential of a drug combination unbiasedly. (5) Finally, we develop data annotation tools to retrieve the pharmacological and molecular profiles of the drugs and cells, facilitating the discovery of mechanisms of action of the most synergistic and effective drug combinations. We provide a new website at www.synergyfinderplus.org to allow users to upload their experimental results and run all the analyses with the R package in the backend. All functions and source codes are freely accessible to academic users (**Figure 1**).

Method

Synergy models for high-order drug combinations

Considering the response of a drug to be measured as a % inhibition y that ranges from 0 to 1, a higher value indicates a better efficacy. For a combination that involves n drugs, the observed combination response is denoted as y_c, whereas the observed monotherapy response of its constituent drugs is $y_{i,i=1,\dots n}$. The expected combination response is determined by the assumption of non-interaction. Currently there are four major synergy reference models [16,17]. For the HSA model, the expected response is the highest monotherapy response, that is, $y_{HSA} = \max(y_1, \dots, y_i, \dots, y_n)$. For the BLISS model, the expected response is derived from the probabilistic independence of the monotherapy effects, that is, $y_{BLISS} = 1 - \prod_i (1 - y_i)$. For the LOEWE model, the expected response satisfies $\sum_i (\frac{x_i}{f_i^{-1}(y_{LOEWE})}) = 1$, where x_i is the dose of each constituent drug in the combination, and $f_i^{-1}(y_{LOEWE})$ is the inverse function of the dose-response curve. For the ZIP model, the expected response satisfies $y_{ZIP} = 1 - \prod_i (1 - \hat{y}_i)$, where $\hat{y}_i$ is the predicted dose response of the monotherapy by a monotonically increasing curve fitting model $\hat{y}_i = f_i(x_i)$. For example, a common choice for modeling drug dose-response curves is the four-parameter log-logistic model $f(x) = \frac{E_{min} + E_{max}(x/m)^{\lambda}}{1 + (x/m)^{\lambda}}$, where E_{min}, E_{max}, m, and λ are the minimal and maximal responses, IC50, and slope of the dose-response curve, respectively.

Accordingly, the multi-drug synergy score for the observed combination response y_c can be determined as follows:

$$S_{HSA} = y_c - \max(y_1, \dots, y_i, \dots, y_n). \quad (1)$$

$$S_{BLISS} = y_c - \left(1 - \prod_i (1 - y_i)\right) \tag{2}$$

$$= y_c - (\sum_i y_i - \sum_{i<j} y_i y_j + (-1)^{r+1} \times \sum_{i<j} \prod_j^r y_j + (-1)^{n+1} \times \prod_i y_i)$$

$$S_{LOEWE} = y_c - y_{LOEWE}, s.t. \sum_i (\frac{x_i}{f_i^{-1}(y_{LOEWE})}) = 1 \tag{3}$$

To determine the ZIP-based synergy score, y_c needs to be replaced with the predicted average response $\hat{y}_c$ given by the curve fitting models f_i' to make it comparable to y_{ZIP}:

$$S_{ZIP} = \hat{y}_c - y_{ZIP} = \frac{1}{n} \sum_i f_i'(x_i) - (1 - \prod_i (1 - f_i(x_i)), \tag{4}$$

$$\text{where } Emin(f_i') = f_{-i}'(x_{-i})$$

Namely, $f_i'(x)$ stands for the log-logistic model defined for the combination response at dose x of drug i when the other drugs are present. Furthermore, $Emin$ of $f_i'(x)$ is determined by $f_{-i}'(x_{-i})$, which is the fitted curve of the combination response while drug i is absent. Note that the ZIP model captures the shift of potency for a drug combination in comparison to its monotherapy drugs; therefore, the ZIP model compares the difference between fitted models for the drug combination $f_i'(x_i)$ and for the monotherapy drug $f_i(x_i)$.

Visualization of higher-order synergy and sensitivity scores

The commonly utilized synergy or sensitivity landscape models are not directly applicable for the visualization of higher-order drug combinations. It is non-trivial to determine the coordinates of the multi-dimensional dose vectors in a two-dimensional space where similar dose vectors remain proximal to each other. In this study we propose a dimension reduction technique that is based on multidimensional scaling, similar to the recent application in transforming numerical data into images [18]. For a drug combination in an *n*-dimensional dose space, with its coordinates of $X = (x_1, \ldots x_n)$, we consider their rankings $R = (r_1, \ldots r_n)$. For example, a combination *X = (0.1, 10, 1000)* nM in three-dimensional dose space has a ranking vector R = (1, 3, 6), suggesting that 0.1 nM is the minimum dose in the tested dose range of drug 1, 10 nM is the third smallest dose of drug 2, and 1000 nM is the sixth dosage of drug 3. Euclidean distance is a common choice for quantifying the similarity between two dose ranking vectors. We utilize multi-dimensional scaling to minimize the error of the pairwise distance in a two-dimensional space in which the synergy and sensitivity scores can be visualized as a synergy landscape (**Figure 2**). Using the dose rankings as the input for multi-dimensional scaling can ensure that the resulting two-dimensional coordinates are equally distanced among the neighboring dose conditions, thus making the visualization easier to interpret.

Furthermore, for the case of two-drug combinations, the algorithm can converge to the actual dose rankings, thus preserving the consistency across all orders of combinations. We also develop functions to visualize the synergy and sensitivity scores for a specific dose condition in a grid of bar plots, which can help identify the most synergistic and sensitive areas in the synergy landscape.

Statistical analysis

The statistical significance of synergy can be defined at a single dose or at the whole dose matrix level. Consider a two-drug combination experiment as an example. We assume that the replicates of drug responses are measured independently within each dose in the dose matrix (**Figure 3**). At each dose level, we propose using a bootstrapping approach to determine the confidence intervals of the synergy scores. To determine a bootstrap dose-response matrix, we sample the responses for each dose combination from a normal distribution $N(\mu, \sigma)$, where μ and σ are the mean and standard deviation of the replicates. The synergy scores for the HSA and BLISS models can be derived simply by comparing the bootstrapped responses at the dose combination and at their corresponding monotherapy doses (*i.e.*, using Equations (1) and (2)). However, for the LOEWE and ZIP models [7], bootstrapped synergy scores will be derived using the curve fitting over the whole dose matrix (*i.e.,* Equations (3) and (4)).

Suppose that B bootstrap samples are drawn from the replicates and suppose that $s_1, \ldots, s_B$ are the estimates of the synergy scores determined from these bootstrap samples, with a mean of $\bar{s}$. We determine the bootstrap standard error (SE) as follows:

$$SE = \sqrt{\frac{1}{B-1}\sum_{i=1}^{B}(s_i - \bar{s})^2} \tag{5}$$

The 95% confidence interval for the synergy score is approximately

$$[\bar{s} - 1.96SE, \bar{s} + 1.96SE]$$

At the whole dose matrix level, we provide an empirical P value to assess the significance of the difference between the estimated average synergy score over the whole dose matrix and the expected synergy score of zero under the null hypothesis of non-interaction.

Letting $s_1', \ldots, s_B'$ are the estimates of the average synergy scores over the whole dose matrix from the bootstrap samples, the P value is [19]:

$$p = \exp(-0.717z - 0.416z^2), \text{ where } z = abs(\bar{s}')\Big/\sqrt{\frac{1}{B-1}\sum_{i=1}^{B}(s_i' - \bar{s}')^2}$$

For a scenario in which no replicates are available, the confidence interval cannot be determined at the dose level. However, the *P* value of the average synergy score of the whole dose-response matrix can still be derived by pooling the synergy scores together and then comparing them to zero. For both scenarios, we report the overall *P* values of synergy and the confidence intervals at the dose-specific level when replicates are available. We test the accuracy of the bootstrapping with the simulation (File S1) and show that the *P* values for synergy scores determined from bootstrap are highly correlated with the true *P* values (Figure S1).

Harmonization of multiple synergy models

As the SynergyFinder R package provides four synergy models (*i.e.*, HSA, BLISS, LOEWE, and ZIP) that differ in their null hypotheses of non-interaction, the synergy scores do not necessarily coincide with each other for a given dataset. To provide a more harmonized representation of the synergy score results, we introduce a novel tool to enable a more systematic comparison of the synergy scores. We leverage the implementation of the four models to derive the expected response of non-interaction (*i.e.*, y_{HSA}, y_{BLISS}, y_{LOEWE}, and y_{ZIP}). Because all the expected responses share the same unit as the observed response (*i.e.*, % inhibition), we may utilize a synergy barometer to compare these values. The expected and observed responses for a given combination at a specific dose are positioned on the same scale. With such a tool as the synergy barometer, one can evaluate the degree of synergy of a specific model and easily understand the differences in the results. A strong synergy should be concluded if the observed response goes beyond the expected responses of all the four models.

Prioritization of drug combinations

In addition to synergy, the sensitivity of a drug combination is equally important [8]. Prioritization of drug combinations based on the degree of synergy may only result in an excessive number of false positives, as a clinical endpoint for approving a drug combination is commonly its efficacy rather than the synergy. In cases where a strong synergy is obtained, the drug combination may not necessarily reach sufficient levels of therapeutic efficacy. To capture the sensitivity of a drug combination, we previously developed a combination sensitivity score (CSS) model that calculates the relative inhibition of a drug combination based on the area under the log10 scaled dose-response curves at the IC50 doses of the constituent drugs [8]. The CSS has the same unit as a drug response (*i.e.*, % inhibition), which makes it directly comparable with the synergy scores defined in Equations (1–4). Therefore, we propose a scatter plot of CSS versus synergy score for a set of drug

198 combinations. The proposed sensitivity-synergy (SS) plot significantly aids in the prioritization of
199 drug combinations from a high-throughput experiment.
200
201 **Annotation of drug combinations**
202 Drugs and cell lines in a drug screening dataset are annotated by querying related databases. PubChem
203 CID, Isomeric SMILES, and standard InChIKey for drugs are extracted from PubChem [20].
204 Furthermore, we provide the molecular formula, clinical stage, disease indication, cross references to
205 multiple major chemical compound databases (*i.e.*, ChEMBL [21], ChEBI [22], DrugBank [23],
206 BindingDB [24], PharmGKB [25], Guide to Pharmacology [26], Selleck, and ZINC [27]), and drug
207 target profiles from MICHA [28]. Cell line annotation is obtained from Cellosaurus [29] (Version
208 37.0).

209 ## Results

210 **Analysis of three-drug combinations**
211 To test the algorithms designed for higher-order combinations, we utilize a recent anti-malaria study
212 in which 16 triple artemisinin-based combination therapies have been tested against 15 parasite lines
213 [30]. The study aimed to identify synergistic partner drugs that can be combined with artemisinin to
214 overcome the emergence of malaria resistance. For each triple drug combination, a 10 × 10 × 12
215 multi-dimensional dose matrix was constructed, resulting in n = 1200 viability values. The datasets
216 are obtained from the NCATS data portal at https://tripod.nih.gov/matrix-client/.
217
218 We analyze the triple drug combinations in terms of their sensitivity and synergy and visualize them
219 in surface plots using dimension reduction techniques. **Figure 4** shows a synergistic triple
220 combination, pyronaridine tetraphosphate-piperaquine-darunavir ethanolate (PYR-PQP-DRV)
221 compared to an antagonistic combination, pyronaridine tetraphosphate-piperaquine-lopinavir (PYR-
222 PQP-LPV). Despite similar sensitivities (the average % inhibition is 78.87 for PYR-PQP-DRV and
223 74.97 for PYR-PQP-LPV), their synergy scores differ drastically (the average ZIP score is 12.02 for
224 PYR-PQP-DRV and -13.39 for PYR-PQP-LPV), suggesting more complex behaviors of the drug
225 interactions as the number of drugs increases. Because of the lack of computational tools for
226 analyzing higher-order drug combinations, the authors of the anti-malaria studies provided only the
227 results of pairwise drug combinations, whereas the actual interactions of the three drugs were
228 unexplored. Therefore, our method may provide novel insights that leverage higher-order drug
229 combination design in a more tailored manner.

230

231 **Analysis of two-drug combinations with replicates**

232 We obtain the ONEIL dataset from the DrugComb data portal [31]. The ONEIL data is a pan-cancer
233 drug combination study in which 583 drug combinations were tested across 39 cell lines [32]. Four
234 replicates were produced for each drug combination, making it a representative example dataset to
235 analyze the significance of the synergy scores. In **Figure 5**, we show the selective drug combinations
236 tested in the A2058 cell line (melanoma).

237

238 As shown in Figure 5A, the confidence intervals for the responses and synergy scores vary at
239 different doses, suggesting that not all scores are equally significant. Furthermore, we observe a
240 general trend that a higher response or synergy score tends to have a smaller confidence interval,
241 consistent with the expectation of a typical drug screening experiment. To evaluate the synergy scores
242 more systematically, we provide a synergy barometer for specific dose combinations. For example,
243 for MK-1775 at 0.2 μM combined with niraparib at 2.75 μM, the drug combination response reaches
244 77.84% inhibition, as indicated by the pointer readout on the barometer (Figure 5B, left panel). Such
245 a drug combination response is synergistic, as the expected responses of the HSA, LOEWE, BLISS,
246 and ZIP models are much smaller, shown as the marks on the barometer. In contrast, paclitaxel at
247 0.0016 μM in combination with L-778123 at 2 μM shows a near-zero response (0.29% inhibition),
248 which is much smaller than the expected response among all the synergy models (Figure 5B, right
249 panel); therefore, it is considered strong antagonism. Using the synergy barometer, the actual drug
250 combination response can be directly compared with the expectations of non-interaction among
251 multiple synergy models.

252

253 Furthermore, we also show the CSS and ZIP synergy scores of all the drug combinations tested in
254 the A2058 cell line in an SS plot (Figure 5C, Figure S3). CSS indicates the efficacy of a drug
255 combination, whereas the synergy score indicates the degree of interaction. To prioritize potential
256 drug combinations, it is necessary to identify those with higher CSS and higher synergy scores, that
257 is, the top-right corner of the SS plot.

258

259 **Annotation of mechanisms of action of drug combinations**

260 Recent advances in machine learning and artificial intelligence studies have shown great potential in
261 predictive modeling of drug synergies [33–35]. In these studies, the chemical structure information
262 regarding drugs and the molecular profiles of cells are considered predictive features. However, for
263 most of the drug combination datasets, annotations of drug and cell features are not provided. To

264 facilitate the interpretation of mechanisms of action of drug combinations, it would be beneficial to
265 provide data integration tools for effective annotation and accurate matching to public databases such
266 that more comprehensive features used for synergy predictions can be obtained. Efforts to annotate
267 and harmonize existing drug screening data, such as DrugComb [31,36] and DrugCombDB [37], have
268 significantly contributed to the development of data-driven pharmacological modeling [38–40]. For
269 newly generated drug screening datasets, we develop the SynergyFinder Plus portal as a companion
270 tool for retrieving publicly available information in a more automated fashion. With one click of the
271 button, users can obtain cross-database identifiers of the drugs and cells, as well as additional
272 information, such as the mechanism of action and disease indication (**Figure 6**).
273
274 We show how SynergyFinder Plus annotates a particular drug combination in the ONEIL [32]
275 dataset (5-fluorouracil and vorinostat tested in the A2058 cell line, Figure 6A). After users upload the
276 dataset, SynergyFinder Plus returns two annotation tables describing the cell and drug identities. For
277 example, the cell annotation table includes synonyms, Cellosaurus accession, tissue origin, disease
278 name, and disease NCI Thesaurus ID, whereas the drug annotation table includes InChIKey, isomeric
279 SMILES, molecular formula, clinical trial phase, cross references, and disease indication. In addition,
280 a third table listing primary and potent drug target information is provided. These annotations enable
281 a more systematic exploration of drugs and cell lines tested in a high-throughput experiment (Figure
282 6B). With such a comprehensive collection of identifiers, we believe that the drug combination data
283 can be better integrated with other related data sources for further downstream analysis.

284 Discussion

285 Synergistic and effective drug combinations have long been studied to improve disease management.
286 Testing potential drug combinations using clinically representative cell cultures has become an
287 important tool for detecting signs of drug sensitivity and resistance; in some cases, the results have
288 already provided informative decision-making resources to guide patient stratification in clinical
289 trials [41]. To facilitate the prioritization of drug combinations, informatics tools that enable formal
290 assessment of the degree of synergy as well as sensitivity are highly needed. We provide a major
291 update to the commonly used SynergyFinder R package, which models drug interactions for any
292 higher-order combinations [15]. Higher-order drug combinations have been recently explored in lung
293 cancer [42], colorectal cancer [43], and ovarian cancer [44]. Therefore, it is expected that promising
294 multi-drug combinations will enter formal clinical trials for multiple diseases [45]. Furthermore, we
295 develope a statistical analysis of drug combinations, which is also generally applicable for higher-

order combinations. Importantly, the mathematical models for higher-order combinations have been consistent with those for two-drug combinations; therefore, their scores are readily comparable across different orders. To visualize higher-order combinations, we develop a dimension reduction technique to display the synergy and the sensitivity landscape on a two-dimensional plane.

In addition to providing multiple synergy models, we address the problem of harmonization. With the increasing number of models that have been proposed to characterize drug synergy, there is an increasing need to help users understand the differences between these models such that a user may make an informed interpretation of the drug combination results. We propose the use of a synergy barometer on which all the synergy models can be systematically compared with the observed drug combination response. Despite only including the four major synergy models, HSA, BLISS, LOEWE, and ZIP, we propose the incorporation of other synergy models that can readily have their synergy scores put into the barometer for comparison with existing models. With the synergy barometer, we hope that the drug discovery community may derive a consensus on the best reporting guideline for drug combination data analysis, which has been long sought since the time of the Saariselkä agreement [46] and thereafter [47].

We also provide the formal statistical analysis for replicates of drug combinations. The confidence intervals and *P* values for all four synergy scores are comparable. Even for the cases in which no replicates have been performed, the significance of synergy over all the dose conditions can be provided, thus enabling a statistical evaluation of drug combinations in a primary screen that usually does not provide replicates due to the high number of combinations. Once a drug combination hit passes the initial screening based on the synergy barometer and the statistical significance, we recommend a confirmation screen that involves more doses and replicates to formally evaluate its synergy. If a user would like to study the mechanism of action of the drug combination, we also provide integrative tools to retrieve the chemical structure information and other functional annotations of the drugs, which will help in downstream analysis of the drug combinations. For higher-order drug combinations, the interpretation of drug synergies is a challenging problem. We hope that our tool provides the first evidence to pinpoint synergistic higher-order drug combinations.

Finally, we provide new features to harmonize the assessment of synergy and sensitivity. The synergy models that have been developed in the SynergyFinder R package have the same scale as the drug combination sensitivity method that was developed earlier, thus allowing a direct comparison in an SS plot. We recommend that users consider the SS plot when reporting their drug combination

results, as sensitivity is the clinical endpoint for approving any drug or drug combinations. The drug combination sensitivity values also have the same scale as monotherapy drug responses in the unit of percentage inhibition; therefore, we can readily compare a drug combination with a single drug to evaluate their efficacy. With the harmonization of drug combination sensitivity and monotherapy sensitivity, we may provide an integrated data source for developing advanced machine learning approaches for predicting drug sensitivity, such as drug sensitivity datasets deposited in DrugComb [31]. To facilitate the FAIRification (findable, accessible, interoperable, and reusable characteristics) of drug combination analysis, we provide a web server at www.synergyfinderplus.org to allow users to upload the data and obtain the analysis results in the form of reports. In addition, we encourage users to deposit their experimental data in DrugComb at https://drugcomb.org/, currently one of the most comprehensive drug screening databases. With the algorithms and the source code made available through the SynergyFinder R package, we welcome the machine learning community to leverage the harmonized drug combination and monotherapy datasets for more robust and transferable predictions to further facilitate the drug combination discovery [34,48–50].

Availability

The www.synergyfinderplus.org web server is located at the CSC-IT Center for Science in Finland. All the source code for implementing the data analysis and visualization is available as the SynergyFinder R package at https://bioconductor.org/packages/release/bioc/html/synergyfinder.html.

Author CrediT statement

Shuyu Zheng: Conceptualization, Methodology, Software, Writing – Original draft preparation **Wenyu Wang**: Visualization, Writing – Original draft preparation **Jehad Aldahdooh**: Software **Alina Malyutina**: Methodology **Tolou Shadbahr**: Data Curation **Ziaurrehman Tanoli**: Resources **Alberto Pessia**: Methodology, Validation, Writing – Review & Editing **Jing Tang**: Supervision, Conceptualization, Methodology, Software, Writing – Original draft preparation

Acknowledgments

This work is supported by the European Research Council (ERC) starting grant DrugComb (Informatics approaches for the rational selection of personalized cancer drug combinations) [No. 716063], European Commission H2020 EOSC-life (Providing an open collaborative space for digital biology in Europe [No. 824087], and the Academy of Finland grant [No. 317680]. The Sigrid Jusélius Foundation grant. SZ and WW hold a salaried position funded by the University of Helsinki through

the Doctoral Program of Biomedicine (DPBM). WW also receives a personal grant from K. Albin Johanssons stiftelse. AM holds a salaried position funded by the University of Helsinki through the Doctoral Program of Integrative Life Science (ILS), and receives a personal grant from the Cancer Foundation Finland.

Conflict of interests

None declared.

ORCID

0000-0003-0624-8077 (Shuyu Zheng)
0000-0003-2816-9740 (Wenyu Wang)
0000-0002-9947-4067 (Jehad Aldahdooh)
0000-0001-8220-5859 (Alina Malyutina)
0000-0001-5715-5933 (Tolou Shadbahr)
0000-0003-2435-9862 (Ziaurrehman Tanoli)
0000-0001-8607-9191 (Alberto Pessia)
0000-0001-7480-7710 (Jing Tang)

References

[1] Vasan N, Baselga J, Hyman DM. A view on drug resistance in cancer. Nature 2019;575:299–309.
[2] Nathan C. Resisting antimicrobial resistance. Nat Rev Microbiol 2020;18:259–60.
[3] Boshuizen J, Peeper DS. Rational cancer treatment combinations: an urgent clinical need. Mol Cell 2020;78:1002–18.
[4] Berenbaum MC. What is synergy? Pharmacol Rev 1989;41:93–141.
[5] Loewe S. The problem of synergism and antagonism of combined drugs. Arzneimittelforschung 1953;3:285–90.
[6] Bliss CI. The toxicity of poisons applied jointly. Ann App Biol 1939;26:585–615.
[7] Yadav B, Wennerberg K, Aittokallio T, Tang J. Searching for drug synergy in complex dose–response landscapes using an interaction potency model. Comput Struct Biotechnol J 2015;13:504–13.
[8] Malyutina A, Majumder MM, Wang W, Pessia A, Heckman CA, Tang J. Drug combination sensitivity scoring facilitates the discovery of synergistic and efficacious drug combinations in cancer. PLoS Computat Biol 2019;15:e1006752.
[9] Chou T-C. Drug combination studies and their synergy quantification using the chou-talalay method. Cancer Res 2010;70:440–6.
[10] Twarog NR, Stewart E, Hammill CV, Shelat AA. BRAID: A unifying paradigm for the analysis of combined drug action. Sci Rep 2016;6:25523.

396 [11] Di Veroli GY, Fornari C, Wang D, Mollard S, Bramhall JL, Richards FM, et al. Combenefit:
397 an interactive platform for the analysis and visualization of drug combinations. Bioinformatics
398 2016;32:2866–8.
399 [12] Ianevski A, He L, Aittokallio T, Tang J. SynergyFinder: a web application for analyzing drug
400 combination dose-response matrix data. Bioinformatics 2017;33:2413–5.
401 [13] Wooten DJ, Albert R. Synergy: a Python library for calculating, analyzing, and visualizing
402 drug combination synergy. Bioinformatics 2020;37:1473–4.
403 [14] Ianevski A, Giri AK, Aittokallio T. SynergyFinder 2.0: visual analytics of multi-drug
404 combination synergies. Nucleic Acids Res 2020;48:W488–93.
405 [15] Cokol-Cakmak M, Cetiner S, Erdem N, Bakan F, Cokol M. Guided screen for synergistic
406 three-drug combinations. PLoS One 2020;15:e0235929.
407 [16] Meyer CT, Wooten DJ, Lopez CF, Quaranta V. Charting the fragmented landscape of drug
408 synergy. Trends Pharmacol Sci 2020;41:266–80.
409 [17] Vlot AHC, Aniceto N, Menden MP, Ulrich-Merzenich G, Bender A. Applying synergy
410 metrics to combination screening data: agreements, disagreements and pitfalls. Drug Discov
411 Today 2019;24:2286–98.
412 [18] Bazgir O, Zhang R, Dhruba SR, Rahman R, Ghosh S, Pal R. Representation of features as
413 images with neighborhood dependencies for compatibility with convolutional neural networks.
414 Nat Commun 2020;11:4391.
415 [19] Altman DG, Bland JM. How to obtain the *P* value from a confidence interval. BMJ
416 2011;343:d2304.
417 [20] Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, et al. PubChem in 2021: new data content
418 and improved web interfaces. Nucleic Acids Res 2021;49:D1388–95.
419 [21] Mendez D, Gaulton A, Bento AP, Chambers J, De Veij M, Félix E, et al. ChEMBL: towards
420 direct deposition of bioassay data. Nucleic Acids Res 2019;47:D930–40.
421 [22] Hastings J, Owen G, Dekker A, Ennis M, Kale N, Muthukrishnan V, et al. ChEBI in 2016:
422 Improved services and an expanding collection of metabolites. Nucleic Acids Res
423 2016;44:D1214–9.
424 [23] Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a major
425 update to the DrugBank database for 2018. Nucleic Acids Res 2018;46:D1074–82.
426 [24] Gilson MK, Liu T, Baitaluk M, Nicola G, Hwang L, Chong J. BindingDB in 2015: a public
427 database for medicinal chemistry, computational chemistry and systems pharmacology.
428 Nucleic Acids Res 2016;44:D1045–53.
429 [25] Whirl-Carrillo M, Huddart R, Gong L, Sangkuhl K, Thorn CF, Whaley R, et al. An evidence-
430 based framework for evaluating pharmacogenomics knowledge for personalized medicine.
431 Clin Pharmacol Ther 2021;110:563–72.
432 [26] Armstrong JF, Faccenda E, Harding SD, Pawson AJ, Southan C, Sharman JL, et al. The
433 IUPHAR/BPS Guide to PHARMACOLOGY in 2020: extending immunopharmacology
434 content and introducing the IUPHAR/MMV Guide to MALARIA PHARMACOLOGY.
435 Nucleic Acids Res 2020;48:D1006–21.
436 [27] Sterling T, Irwin JJ. ZINC 15 – ligand discovery for everyone. J Chem Inf Model
437 2015;55:2324–37.
438 [28] Tanoli Z, Aldahdooh J, Alam F, Wang Y, Seemab U, Fratelli M, et al. Minimal information
439 for chemosensitivity assays (MICHA): a next-generation pipeline to enable the FAIRification
440 of drug screening experiments. Brief Bioinform 2021;bbab350.
441 [29] Bairoch A. The Cellosaurus, a cell-line knowledge resource. J Biomol Tech 2018;29:25–38.
442 [30] Ansbro MR, Itkin Z, Chen L, Zahoranszky-Kohalmi G, Amaratunga C, Miotto O, et al.
443 Modulation of triple artemisinin-based combination therapy pharmacodynamics by
444 plasmodium falciparum genotype. ACS Pharmacol Transl Sci 2020;3:1144–57.

[31] Zagidullin B, Aldahdooh J, Zheng S, Wang W, Wang Y, Saad J, et al. DrugComb: an integrative cancer drug combination data portal. Nucleic Acids Res 2019;47:W43–51.
[32] O'Neil J, Benita Y, Feldman I, Chenard M, Roberts B, Liu Y, et al. An unbiased oncology compound screen to identify novel combination strategies. Mol Cancer Ther 2016;15:1155–62.
[33] Preuer K, Lewis RPI, Hochreiter S, Bender A, Bulusu KC, Klambauer G. DeepSynergy: predicting anti-cancer drug synergy with deep learning. Bioinformatics 2018;34:1538–46.
[34] Menden MP, Wang D, Mason MJ, Szalai B, Bulusu KC, Guan Y, et al. Community assessment to advance computational prediction of cancer drug combinations in a pharmacogenomic screen. Nat Commun 2019;10:2674.
[35] Li H, Li T, Quang D, Guan Y. Network propagation predicts drug synergy in cancers. Cancer Res 2018;78:5446–57.
[36] Zheng S, Aldahdooh J, Shadbahr T, Wang Y, Aldahdooh D, Bao J, et al. DrugComb update: a more comprehensive drug sensitivity data repository and analysis portal. Nucleic Acids Res 2021;49:W174–84.
[37] Liu H, Zhang W, Zou B, Wang J, Deng Y, Deng L. DrugCombDB: a comprehensive database of drug combinations toward the discovery of combinatorial therapy. Nucleic Acids Res 2020;48:D871–81.
[38] Douglass EF, Allaway RJ, Szalai B, Wang W, Tian T, Fernández-Torras A, et al. A community challenge for pancancer drug mechanism of action inference from perturbational profile data. BioRxiv 2020:2020.12.21.423514.
[39] Zhang B, Tang C, Yao Y, Chen X, Zhou C, Wei Z, et al. The tumor therapy landscape of synthetic lethality. Nat Commun 2021;12:1275.
[40] Shah K, Ahmed M, Kazi JU. The Aurora kinase/β-catenin axis contributes to dexamethasone resistance in leukemia. NPJ Precis Oncol 2021;5:13.
[41] Jones PA, Ohtani H, Chakravarthy A, De Carvalho DD. Epigenetic therapy in immune-oncology. Nat Rev Cancer 2019;19:151–61.
[42] Fernandes Neto JM, Nadal E, Bosdriesz E, Ooft SN, Farre L, McLean C, et al. Multiple low dose therapy as an effective strategy to treat EGFR inhibitor-resistant NSCLC tumours. Nat Commun 2020;11:3157.
[43] Ozkan-Dagliyan I, Diehl JN, George SD, Schaefer A, Papke B, Klotz-Noack K, et al. Low-dose vertical inhibition of the RAF-MEK-ERK cascade causes apoptotic death of *KRAS* mutant cancers. Cell Rep 2020;31:107764.
[44] Caumanns JJ, van Wijngaarden A, Kol A, Meersma GJ, Jalving M, Bernards R, et al. Low-dose triple drug combination targeting the PI3K/AKT/mTOR pathway and the MAPK pathway is an effective approach in ovarian clear cell carcinoma. Cancer Lett 2019;461:102–11.
[45] Van Cutsem E, Huijberts S, Grothey A, Yaeger R, Cuyle P-J, Elez E, et al. Binimetinib, encorafenib, and cetuximab triplet therapy for patients with *BRAF* V600E–mutant metastatic colorectal cancer: safety lead-in results from the phase III BEACON colorectal cancer study. J Clin Oncol 2019;37:1460–9.
[46] Greco W, Unkelbach H, Pöch G, Sühnel J, Kundi M, Bödeker W. Consensus on concepts and terminology for combined-action assessment: the Saariselkä agreement. Arch Complex Environ Stud 1992;4:65–9.
[47] Tang J, Wennerberg K, Aittokallio T. What is synergy? The Saariselkä agreement revisited. Front Pharmacol 2015;6:181.
[48] Wood K, Nishida S, Sontag ED, Cluzel P. Mechanism-independent method for predicting response to multidrug combinations in bacteria. Proc Natl Acad Sci U S A. 2012;109:12254–9.

[49] Adam G, Rampášek L, Safikhani Z, Smirnov P, Haibe-Kains B, Goldenberg A. Machine learning approaches to drug response prediction: challenges and recent progress. NPJ Precis Oncol 2020;4:19.
[50] Kim Y, Zheng S, Tang J, Jim Zheng W, Li Z, Jiang X. Anticancer drug synergy prediction in understudied tissues using transfer learning. J Am Med Inform Assoc 2021;28:42–51.

Figure legends

Figure 1 Summary of the SynergyFinder Plus workflow

We highlight the new features including: (1) the ability to analyze drug combinations of more than two drugs, (2) tailored statistical testing to account for uncertainty for data with replicates, (3) harmonized visualization of multiple synergy score metrics, (4) visualization of the relationship between sensitivity and synergy scores, and (5) a data annotation web server to query cell line and drug information.

Figure 2 Dimension reduction for the visualization of high-order drug combinations

In this example, the two-dimensional coordinates of each three-drug dose combination are determined by multidimensional scaling based on the similarity of their dose ranges, after which the synergy scores and sensitivity scores can be visualized on the map. Each dot in the map is annotated with its associated dose rankings for the three drugs $(r_1; r_2; r_3)$. For example, a dot with a label 4;1;2 shows a combination involves 4th dose of drug 1, 1st dose of drug 2, and 2nd dose of drug 3.

Figure 3 Statistical testing of synergy scores

Dose-specific confidence intervals and *P* values for the average synergy score can be derived using bootstrapping of the dose-response matrix.

Figure 4 Example of surface plots for higher-order drug combinations

A. The dose-response map for a synergistic drug combination (pyronaridine tetraphosphate-piperaquine-darunavir ethanolate, left panel) compared to an antagonistic drug combination (pyronaridine tetraphosphate-piperaquine-lopinavir, right panel). **B**. The ZIP synergy score map for the same drug combinations. For surface plots of other synergy scores, including HSA, BLISS, and LOEWE, see Figure S2.

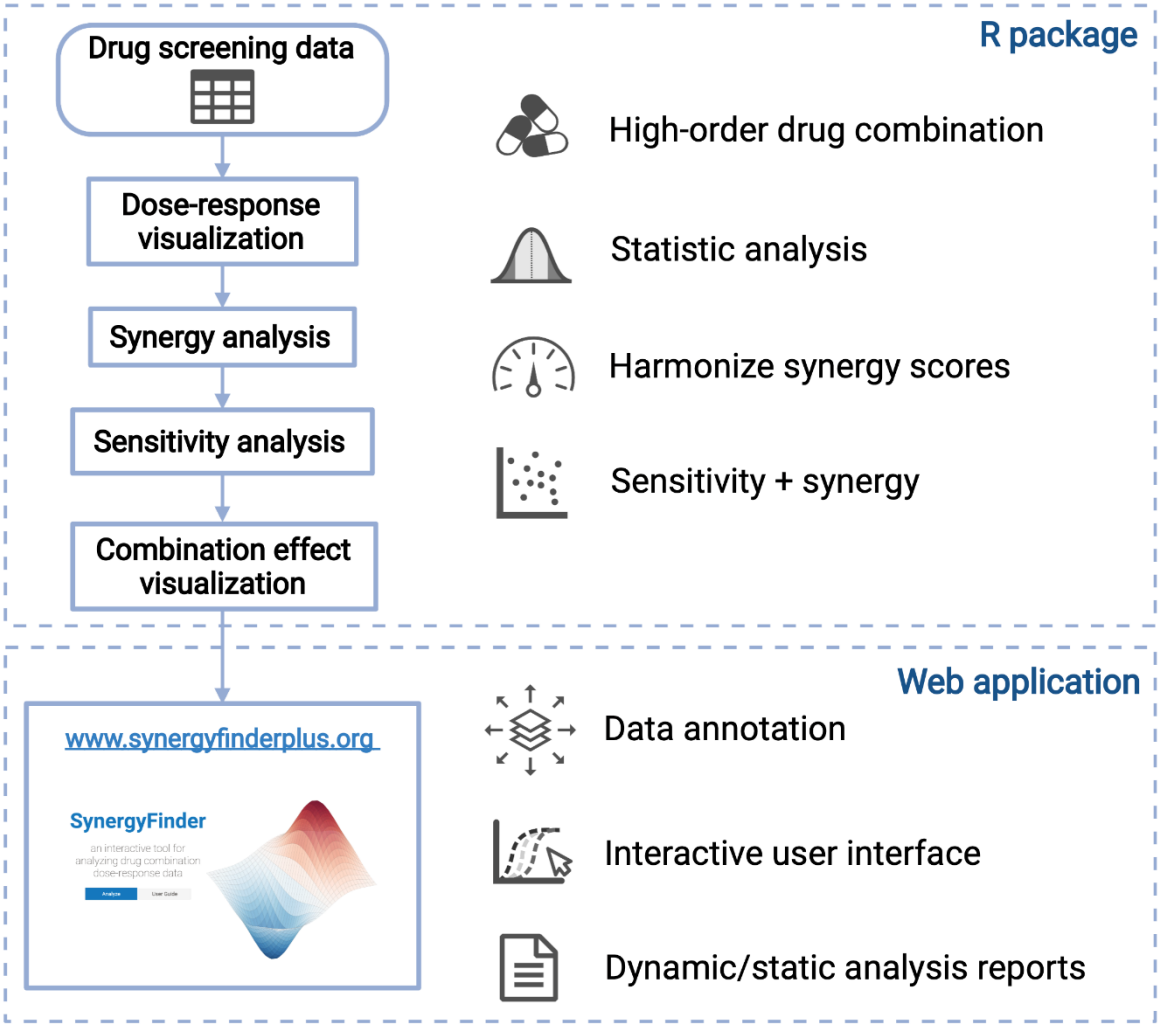
Figure 5 Example of visualizations for two-drug combinations with replicates

527 **A**. An example of the confidence interval plots of a synergistic drug pair (MK-1775 and niraparib;
528 upper panel) and an antagonistic drug pair (paclitaxel and L-778123; lower panel). **B**. The synergy
529 barometer for a given dose combination (left panel: MK-1775 at 0.2 μM in combination with
530 niraparib at 2.75 μM; right panel: paclitaxel at 0.0016 μM in combination with L-778123 at 2 μM).
531 **C**. The synergy-sensitivity plot for drug pairs tested in the A2058 cell line.
532
533 **Figure 6 Example of mechanisms of action annotation for drug combination**
534 **A**. An example of a drug screening annotation table. The tested drug combination, 5-fluorouracil and
535 vorinostat, is further annotated with their chemical and target information. Cell line information for
536 the A2058 is also provided, including synonyms, tissue, and disease types. **B**. Summary of drug
537 clinical stage, drug target protein class, and cell line tissue origin for the ONEIL dataset.

538 ## Supplementary material

539 **Figure S1 Accuracy of *P* values for synergy scores by bootstrapping**
540 **A.** The expected monotherapy responses are defined by four-parameter log-logistic functions and the
541 expected combinational responses are defined by the non-interaction models. Based on the expected
542 values, the distribution of the test statistic under null hypothesis can be determined by simulations.
543 **B**. The scatter plots for comparing the true *P* values and bootstrapping *P* values for 200 samples.
544
545 **Figure S2 Two-dimensional surface plots of synergy scores**
546 **A.** HSA, **B.** BLISS, and **C.** LOEWE for a synergistic three-drug combination (pyronaridine
547 tetraphosphate-piperaquine-darunavir ethanolate, left panel) compared to an antagonistic drug
548 combination (pyronaridine tetraphosphate-piperaquine-lopinavir, right panel).
549
550 **Figure S3 Synergy-sensitivity plots for HSA, BLISS, and LOEWE scores in relation to the**
551 **combination sensitivity score for the drug pairs tested in the A2058 cell line**
552
553 **Table S1 The key differences in the mathematical modeling between SynergyFinderPlus and**
554 **SynergyFinder2 for three-drug combinations**
555
556 **File S1 The simulation code for testing accuracy of *P* values for synergy scores by bootstrapping**
557
558
559
560

561 **Figure 1**



562
563
564
565
566
567
568
569
570
571
572
573
574
575

Figure 2

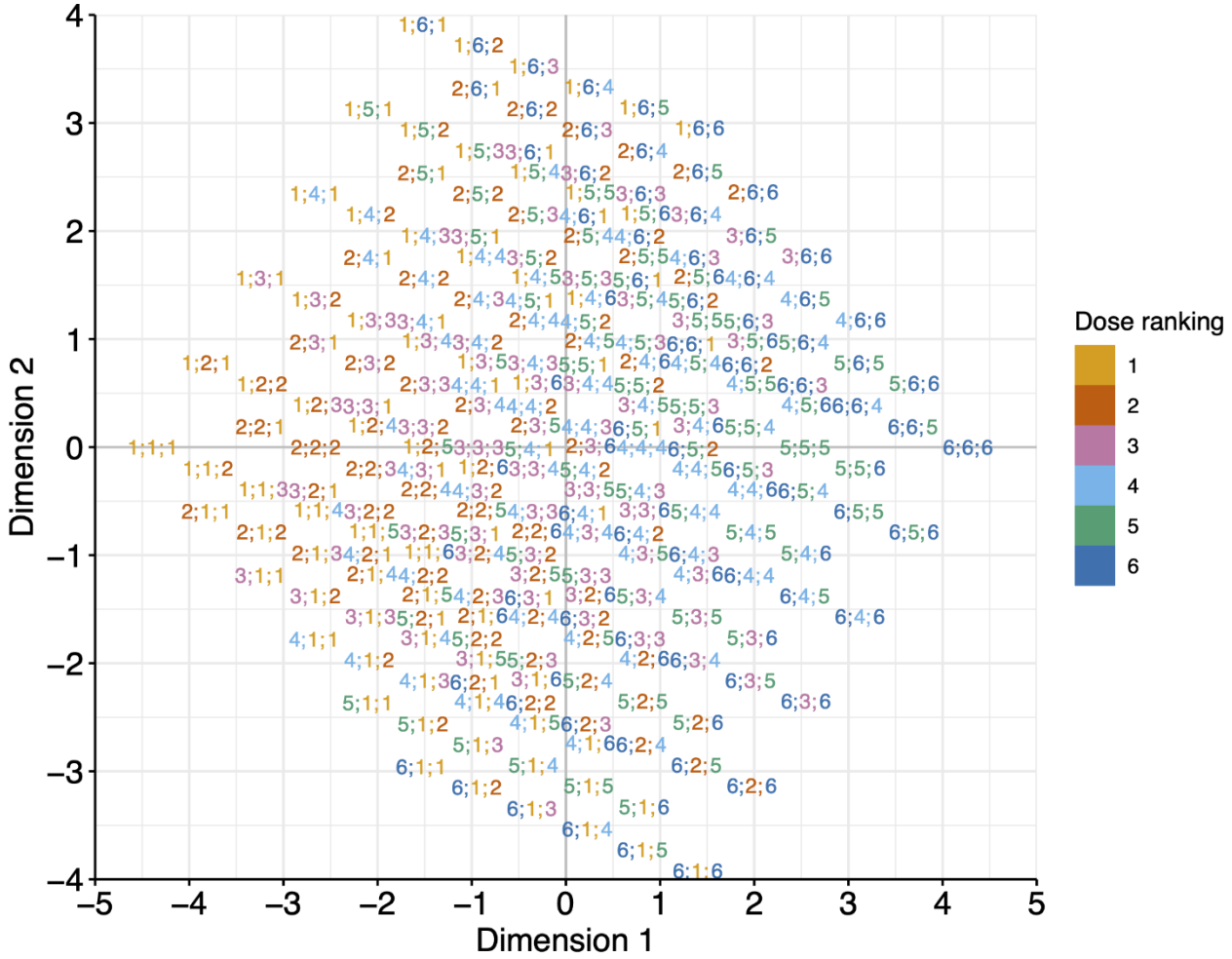
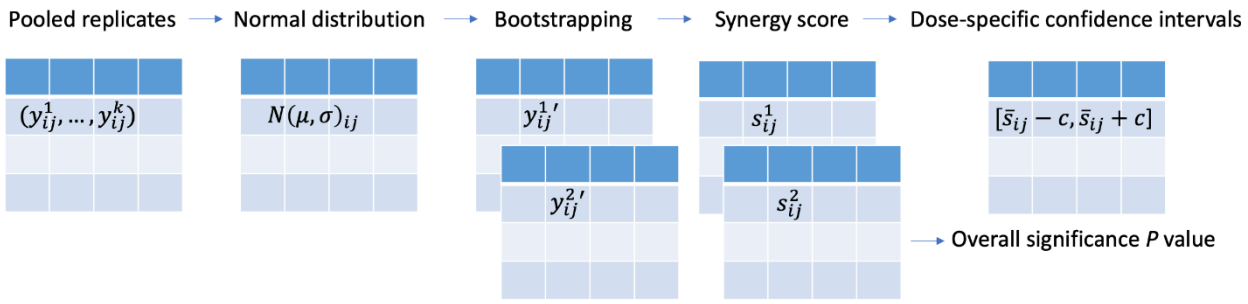
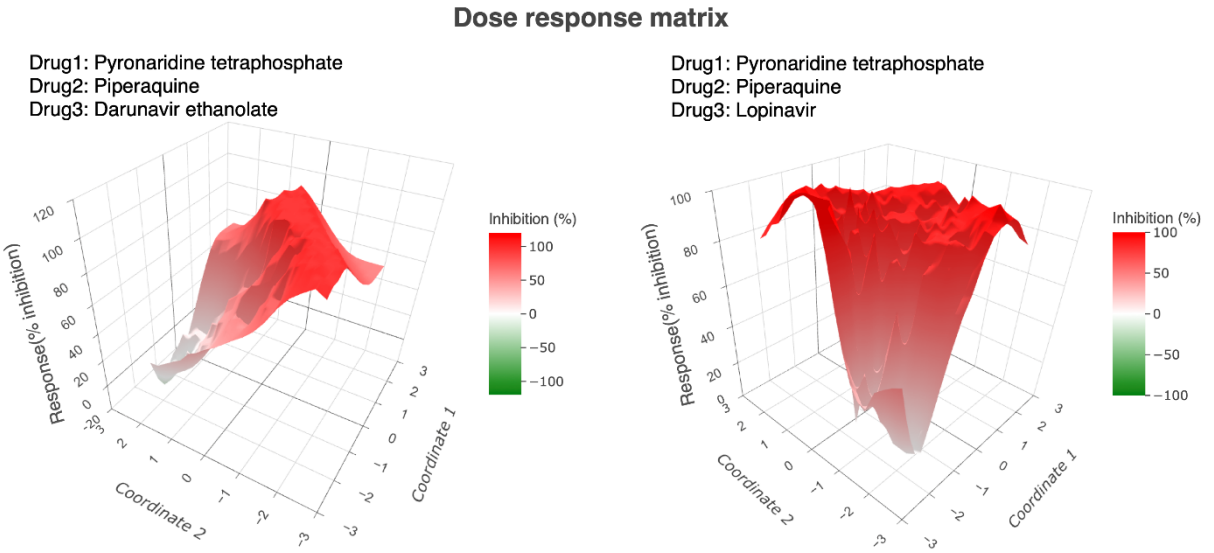


Figure 3

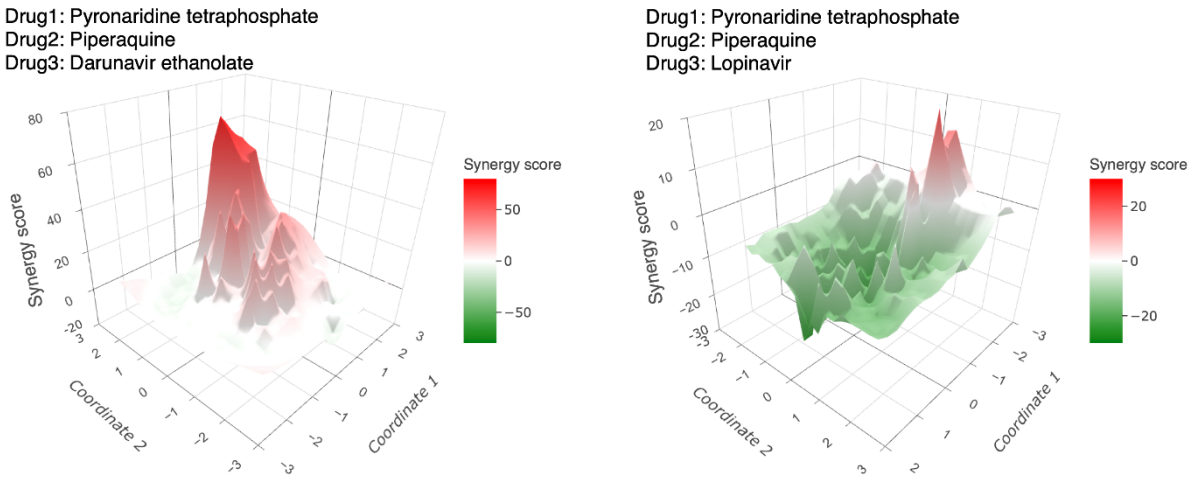

Pooled replicates
Normal distribution
Bootstrapping
Synergy score
Dose-specific confidence intervals
$(y_{ij}^1, \ldots, y_{ij}^k)$
$N(\mu, \sigma)_{ij}$
$y_{ij}^{1\prime}$
$y_{ij}^{2\prime}$
s_{ij}^1
s_{ij}^2
$[\bar{s}_{ij} - c, \bar{s}_{ij} + c]$
Overall significance *P* value

582 **Figure 4**

A



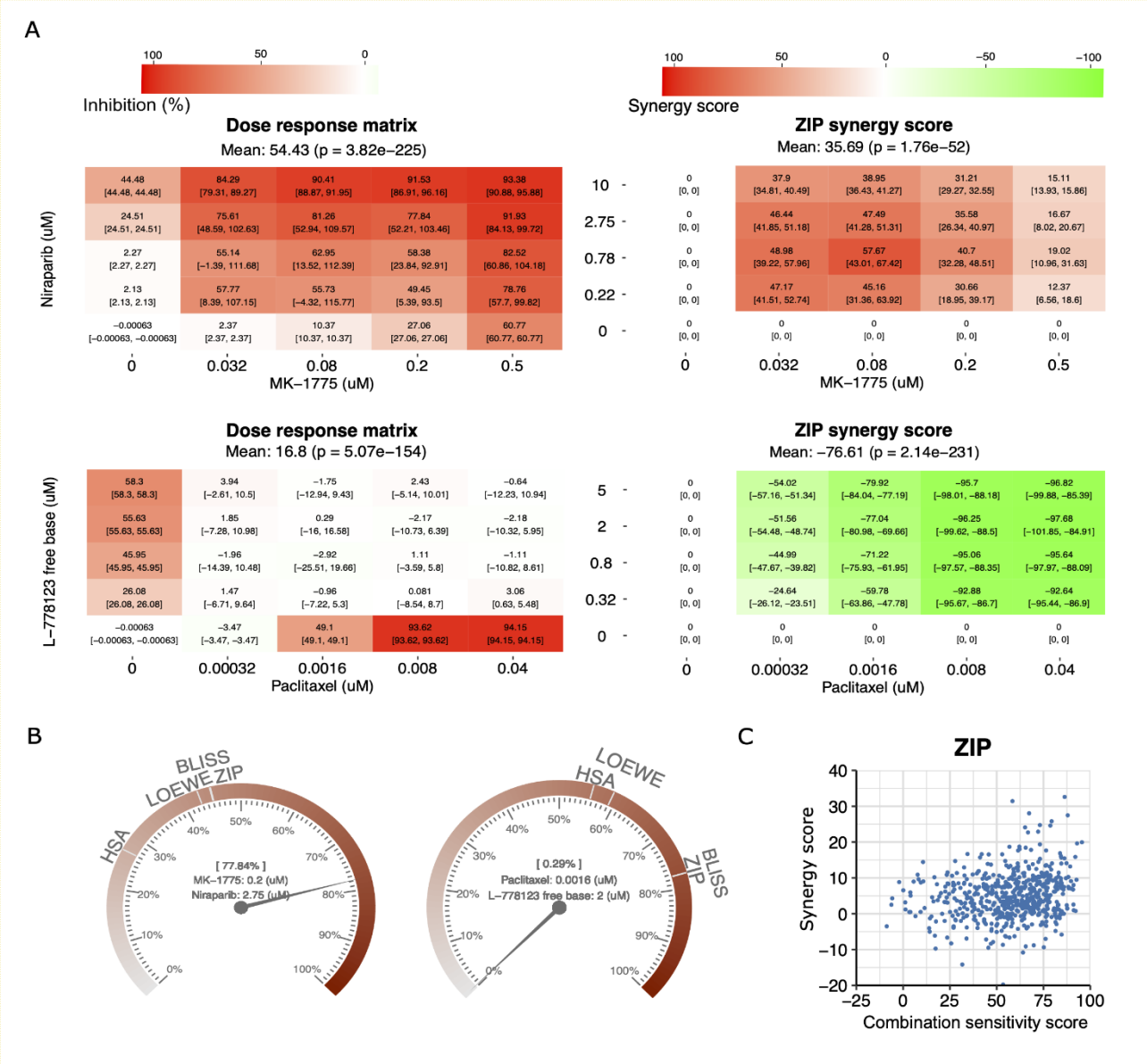
B



583

584

585 **Figure 5**



586

587

588

Figure 6

A

Screening dataset

cell_line_name	drug1	conc1	drug2	conc2	conc_unit1	response	block_id
A2058	5-Fluorouracil	0	Vorinostat	0	um	0	1
A2058	5-Fluorouracil	0	Vorinostat	0.0925	um	-6.16	1
A2058	5-Fluorouracil	0	Vorinostat	0.325	um	-6.13	1
A2058	5-Fluorouracil	0	Vorinostat	1.15	um	12.36	1
A2058	5-Fluorouracil	0	Vorinostat	4	um	82.63	1
A2058	5-Fluorouracil	0.35	Vorinostat	0	um	-1.75	1
A2058	5-Fluorouracil	0.35	Vorinostat	0.0925	um	-10.41	1
A2058	5-Fluorouracil	0.35	Vorinostat	0.325	um	-8.65	1
A2058	5-Fluorouracil	0.35	Vorinostat	1.15	um	30.96	1
A2058	5-Fluorouracil	0.35	Vorinostat	4	um	88.03	1

Showing 1 to 10 of 100 entries Previous 1 2 3 4 5 ... 10 Next

Annotate

Cell info table

Cell Name	Synonyms	Cellosaurus Accession	Tissue	Disease Name	Disease NCIt ID
A2058	A2058; A 2058; A-2058	CVCL_1059	skin	Amelanotic melanoma	C3802

Showing 1 to 1 of 1 entries Previous 1 Next

Drug info table

Drug Name	InChIKey	Isomeric SMILES	Molecular Formula	Max Phase	Cross Reference	Disease Indication
5-Fluorouracil	GHASVSINZRGABV-UHFFFAOYSA-N	C1=C(C(=O)NC(=O)N1)F	C4H3FN2O2	4	PubChem:3385 ChEMBL:CHEMBL185 DrugBank:DB00544 ChEBI:46345 PharmGKB:PA128406956 PubChem:3385 Guide to Pharmacology:4789 Selleck:Adrucil(Fluorouracil) Zinc:ZINC38212689	pancreatic carcinoma, bladder carcinoma, ga carcinoma, biliary tract neoplasm, squamous c carcinoma, Gastrointes and neck malignant ne adenocarcinoma, prolif neoplasm, actinic kera neoplasm, uterine neop neoplasm, nasopharyn carcinoma, Merkel cell gallbladder neoplasm, neoplasm
Vorinostat	WAEXFXRVDQXREF-UHFFFAOYSA-N	C1=CC=C(C=C1)NC(=O)CCCCCCC(=O)NO	C14H20N2O3	4	PubChem:5311 Selleck:Vorinostat-(SAHA) DrugBank:DB02546 Guide to Pharmacology:6852 ChEBI:45716 Zinc:ZINC01543873 PubChem:5311 ChEMBL:CHEMBL98	multiple myeloma, uret infection, Cutaneous T- lymphoblastic leukemi marginal zone B-cell ly Hodgkins lymphoma, e carcinoma, precursor T astrocytoma, lymphoid myelodysplastic syndr adenocarcinoma, ovari gastric carcinoma, Mal myelogenous leukemia ovarian carcinoma, neu acute promyelocytic le carcinoma, chronic lym schizophrenia, MALT ly

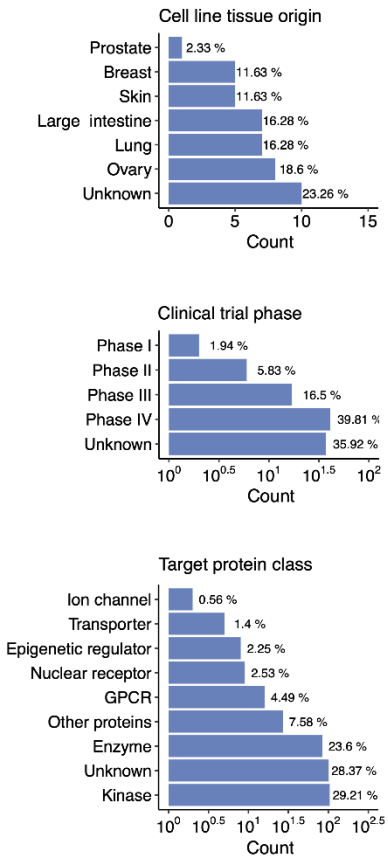
Showing 1 to 2 of 2 entries Previous 1 Next

Drug target info table

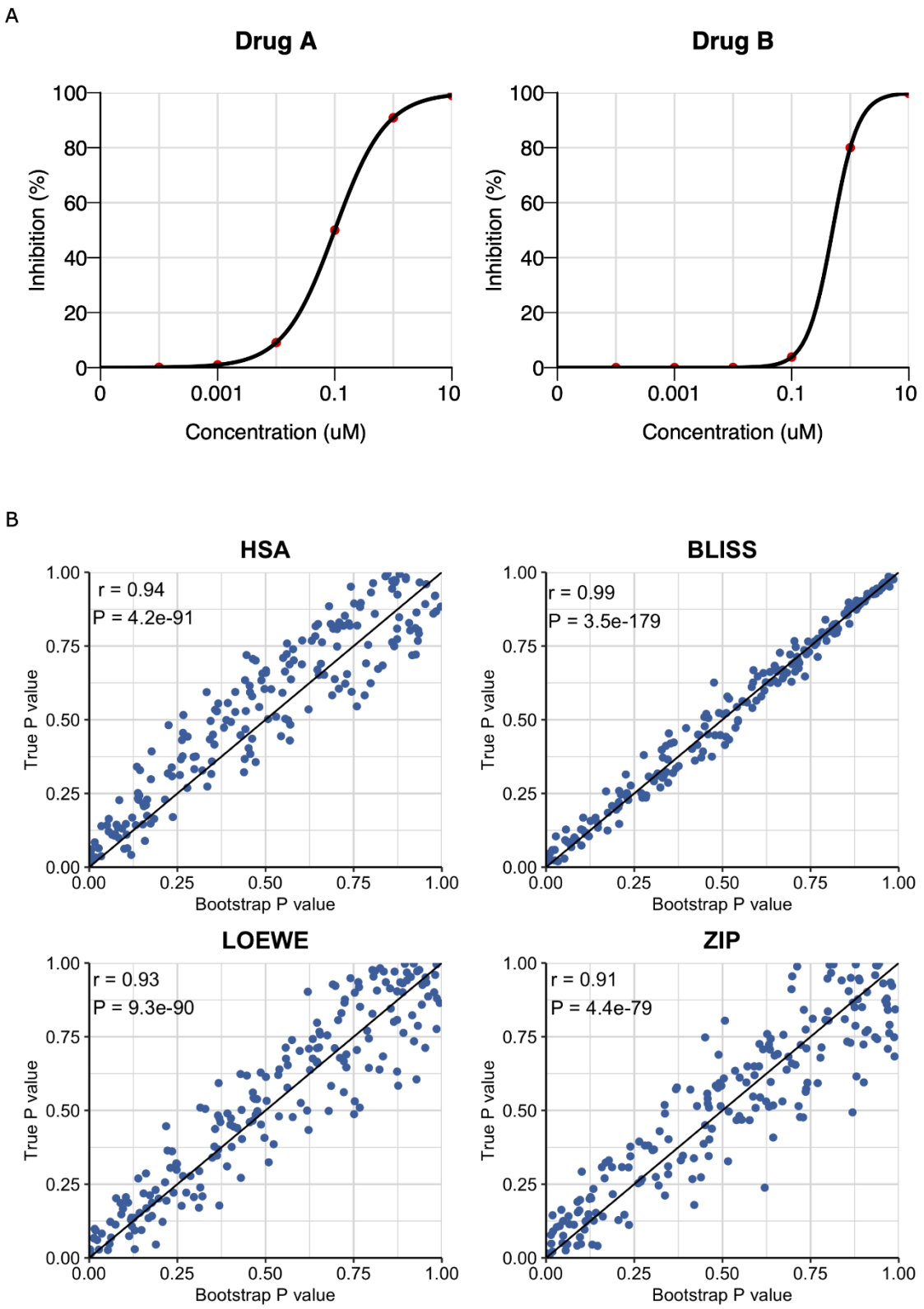
Drug Name	InChIKey	Primary Target Name	Primary Target ID	Potent Target Name*
5-Fluorouracil	GHASVSINZRGABV-UHFFFAOYSA-N	TYMS	P04818	,APEX1,EHMT2,LMNA,NR3C1,THPO
Vorinostat	WAEXFXRVDQXREF-UHFFFAOYSA-N	HDAC1,HDAC3,HDAC6,HDAC2	Q13547,O15379,Q9UBN7,Q92769	,HDAC1,HDAC10,HDAC11,HDAC2,HDAC3,

Showing 1 to 2 of 2 entries Previous 1 Next

B



592 **Figure S1**



593

594

Figure S2

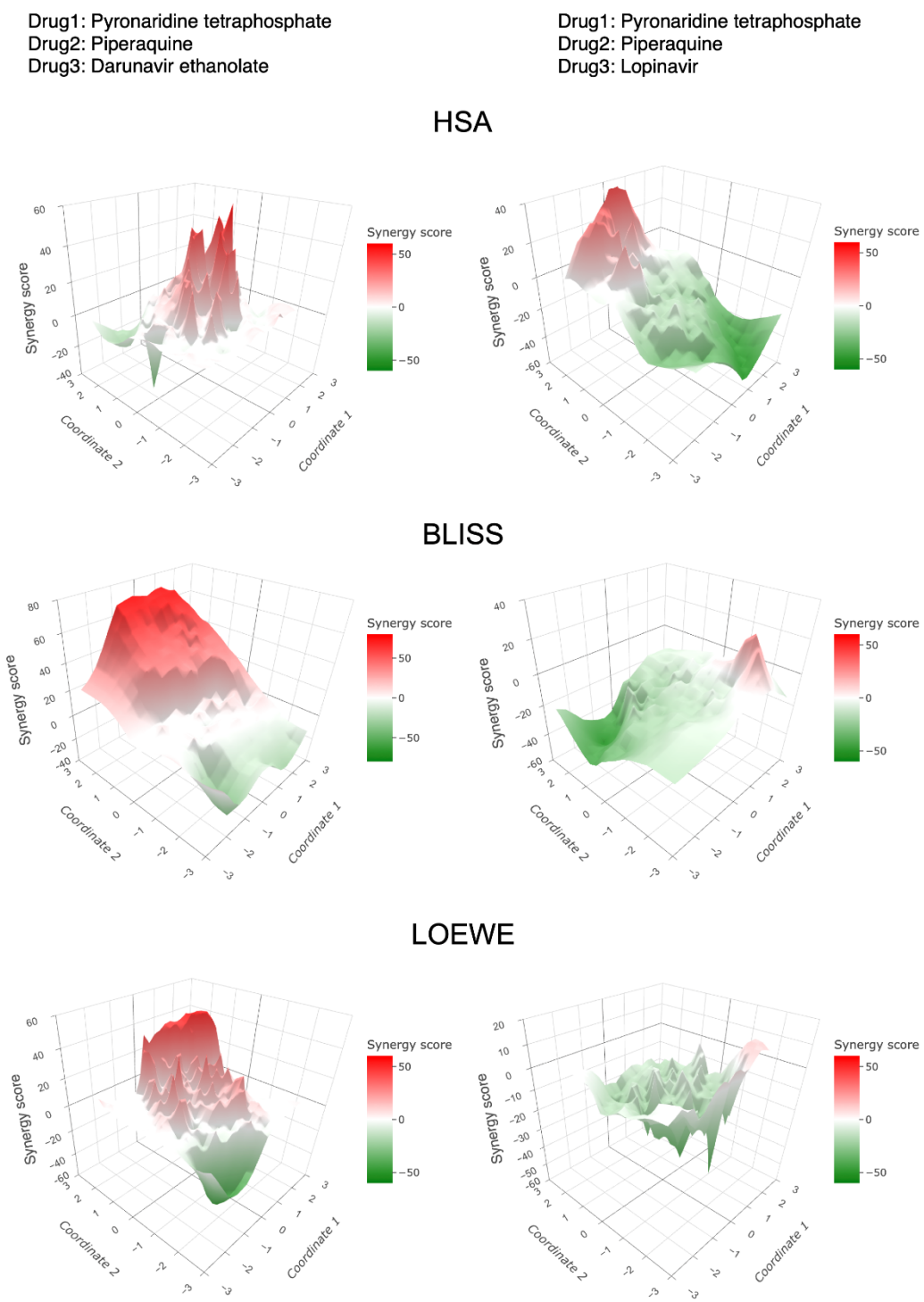

Drug1: Pyronaridine tetraphosphate
Drug2: Piperaquine
Drug3: Darunavir ethanolate
Drug1: Pyronaridine tetraphosphate
Drug2: Piperaquine
Drug3: Lopinavir
HSA
BLISS
LOEWE
Synergy score
Coordinate 1
Coordinate 2

Figure S3

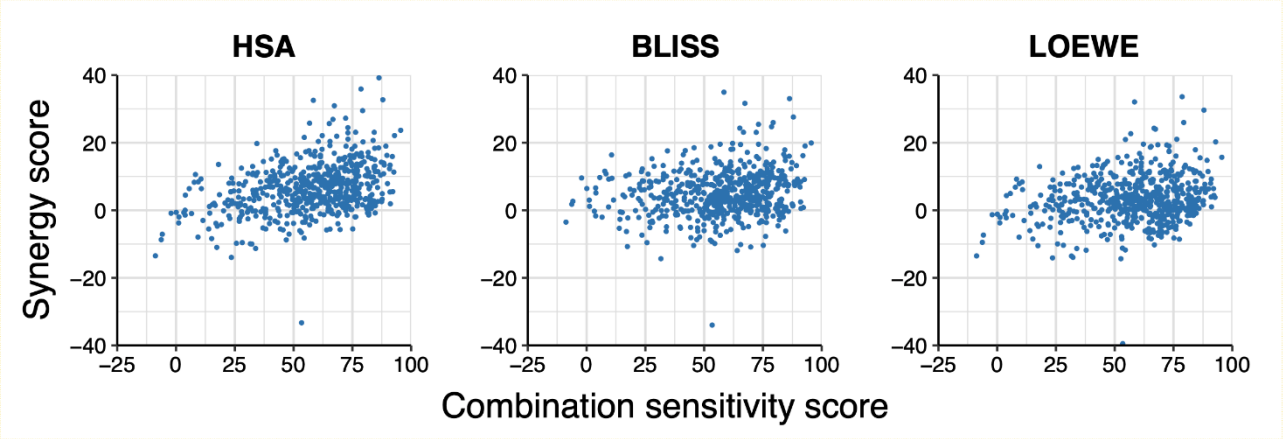

HSA
BLISS
LOEWE
Synergy score
Combination sensitivity score
40
20
0
−20
−40
−25
0
25
50
75
100

Table S1 The key differences in the mathematical modeling between SynergyFinderPlus and SynergyFinder2 for three-drug combinations

	SynergyFinderPlus	SynergyFinder2
BLISS	$S_{BLISS} = E_{A,B,C} - (E_A + E_B + E_C - E_A E_B - E_A E_C - E_B E_C + E_A E_B E_C)$	$S_{BLISS} = E_{A,B,C} - (E_A + E_B + E_C - E_A E_B - E_A E_C - E_B E_C - E_A E_B E_C)$
LOEWE	$S_{LOEWE} = E_{A,B,C} - E_{LOEWE}$, $s.t. \sum_{i \in \{A,B,C\}} (\frac{x_i}{f_i^{-1}(E_{LOEWE})}) = 1$	$S_{LOEWE} = \frac{a}{A} + \frac{b}{B} + \frac{c}{C}$
ZIP	$S_{ZIP} = \hat{E}_{A,B,C} - (\hat{E}_A + \hat{E}_B + \hat{E}_C - \hat{E}_A \hat{E}_B - \hat{E}_A \hat{E}_C - \hat{E}_B \hat{E}_C + \hat{E}_A \hat{E}_B \hat{E}_C)$ $\hat{E}_{A,B,C} = \frac{1}{3}(\frac{\hat{E}_{-A} + \left(\frac{x_A}{\hat{m}_A}\right)^{\hat{\lambda}_A}}{1 + \left(\frac{x_A}{\hat{m}_A}\right)^{\hat{\lambda}_A}} + \frac{\hat{E}_{-B} + \left(\frac{x_B}{\hat{m}_B}\right)^{\hat{\lambda}_B}}{1 + \left(\frac{x_B}{\hat{m}_B}\right)^{\hat{\lambda}_B}} + \frac{\hat{E}_{-C} + \left(\frac{x_C}{\hat{m}_C}\right)^{\hat{\lambda}_C}}{1 + \left(\frac{x_C}{\hat{m}_C}\right)^{\hat{\lambda}_C}})$	$S_{ZIP} = E_{A,B,C} - (\hat{E}_A + \hat{E}_B + \hat{E}_C - \hat{E}_A \hat{E}_B - \hat{E}_A \hat{E}_C - \hat{E}_B \hat{E}_C - \hat{E}_A \hat{E}_B \hat{E}_C)$