

# DrugCombDB: a comprehensive database of drug combinations toward the discovery of combinatorial therapy

Hui Liu<sup>1</sup>, Wenhao Zhang<sup>1</sup>, Bo Zou<sup>2</sup>, Jinxian Wang<sup>2</sup>, Yuanyuan Deng<sup>2</sup> and Lei Deng<sup>2,3,\*</sup>

<sup>1</sup>Lab of Information Management, Changzhou University, Changzhou 213164, China, <sup>2</sup>School of Computer Science and Engineering, Central South University, Changsha 410075, China and <sup>3</sup>School of Software, Xinjiang University, Urumqi 830008, China

Received August 15, 2019; Revised October 14, 2019; Editorial Decision October 14, 2019; Accepted October 17, 2019

## ABSTRACT

Drug combinations have demonstrated high efficacy and low adverse side effects compared to single drug administration in cancer therapies and thus have drawn intensive attention from researchers and pharmaceutical enterprises. Due to the rapid development of high-throughput screening (HTS), the number of drug combination datasets available has increased tremendously in recent years. Therefore, there is an urgent need for a comprehensive database that is crucial to both experimental and computational screening of synergistic drug combinations. In this paper, we present DrugCombDB, a comprehensive database devoted to the curation of drug combinations from various data sources: (i) HTS assays of drug combinations; (ii) manual curations from the literature; and (iii) FDA Orange Book and external databases. Specifically, DrugCombDB includes 448 555 drug combinations derived from HTS assays, covering 2887 unique drugs and 124 human cancer cell lines. In particular, DrugCombDB has more than 6000 000 quantitative dose responses from which we computed multiple synergy scores to determine the overall synergistic or antagonistic effects of drug combinations. In addition to the combinations extracted from existing databases, we manually curated 457 drug combinations from thousands of PubMed publications. To benefit the further experimental validation and development of computational models, multiple datasets that are ready to train prediction models for classification and regression analysis were constructed and other significant related data were gathered. A website with a user-friendly graphical visualization has been developed for users to access the wealth of data and down-

load prebuilt datasets. Our database is available at <http://drugcombdb.denglab.org/>.

## INTRODUCTION

Although ‘targeted’ drugs have made remarkable advances in the treatment of cancer patients, their clinical benefits are greatly limited due to natural and acquired drug resistance of cancer cells (1). The endogenous mechanism of drug resistance lies in compensatory signal transduction and cross-talk among pathways resulting from long-term evolution (2,3). ‘One-target’ drug treatments often lead to the activation of the compensatory signaling pathway that maintains the growth and survival of tumor cells (4). In contrast, drug combinations have demonstrated great advantages in overcoming drug resistance and improving therapeutic efficacy in cancer therapy and have thus drawn increasing attention from researchers and pharmaceutical enterprises (5,6). There is an emerging trend shifting from single-target to multitarget and combination paradigms in drug discovery (7). However, despite the increasing number of successful drug combinations, most of them are discovered by clinical experience or chance (6,7). Therefore, there is an urgent demand for a rational and systematic methodology to screen cancer-specific and sensitive combinatorial drugs for cancer therapy (8–10). With insight into the endogenous mechanism and pathway interdependencies critical for cancer cell proliferation and survival, we are able to design multiple agents to synergistically inhibit pathogenic pathways (11,12). However, wet-lab experiments for dissecting the cellular mechanism of cascading signal transduction and signaling networks are cost-intensive and time-consuming (13).

Due to the rapid development of high-throughput screening (HTS), it is possible to simultaneously evaluate the sensitivities of drug combinations to hundreds of cancer cell lines (14). As a result, the number of experimental screening datasets available has increased tremendously in recent years, especially the number of dual-agent combinations in-

\*To whom correspondence should be addressed. Tel: +86 18874929663; Email: leideng@csu.edu.cn

volving many FDA-approved drugs (15). These large-scale datasets of drug combinations could greatly benefit both the academic and industrial communities. However, the existing database DCDB (16), which has not been updated since 2014, covers a relatively small number of drug combination annotations extracted from FDA Orange Books and records of clinical trials. The web server DrugComb (released during the preparation of our database) focuses on the computation and visualization of the synergy score of drug combinations (17). Therefore, there is an urgent need for a comprehensive database to collect and integrate the increasing numbers of datasets, which is beneficial to both the experimental and computational screening of drug combinations.

In this paper, we present DrugCombDB, a comprehensive database dedicated to collecting drug combinations from various data sources. Concretely, our combination database covers (i) HTS assays, (ii) manual curations from the PubMed literature, (iii) FDA-approved and investigational combinatorial therapies and (iv) failed drug combinations. The current release of DrugCombDB includes 6055 926 quantitative dose responses from which we computed multiple synergy scores to determine the overall synergistic or antagonistic effects of drug combinations. These synergy scores, determined based on different models such as the Bliss and Loewe independence models, allow us to build quantitative training sets. Particularly, we merely focused on the overall probability of synergistic or antagonistic effect that combinations have, such that the validation of specific dosage with the optimal effect is encouraged in further research. There are also 457 drug combinations that have been manually curated from more than 6000 PubMed publications. In total, DrugCombDB includes 448 555 drug combinations covering 2887 unique drugs and 124 human cancer cell lines. To facilitate the downstream usage of our data resource, we have prepared multiple datasets that are ready for building prediction models for classification and regression analysis. A website with user-friendly graphical visualization has been developed for users to access the wealth of data. Users can input a drug of interest to retrieve associated drug combinations, as well as dose-response landscapes, supporting evidence, drug targets and other useful information.

To the best of our knowledge, DrugCombDB is the first comprehensive database with the largest number of drug combinations to date. We believe that it would greatly facilitate and promote the discovery of novel synergistic drugs for the therapy of complex diseases and cancers. In fact, our database has received extensive attention since our bioRxiv preprint was released in December 2018. We have worked continuously to add more data to DrugCombDB, and the number of experimental dose responses has increased more than 10-fold since the release of the first version of the database.

## DATA SOURCES

### HTS assays

HTS techniques have been widely used to measure the quantitative dose responses of cancer cells to various drug

combinations at different concentrations; therefore, dose-response landscapes were constructed and used to evaluate the combinatorial efficacy (synergy, additivity and antagonism) of drug combinations. For convenience, we introduce the definition of ‘combination test’, which is actually the dose response level (often represented by the  $IC_{50}$  value) of cancer cells to a certain drug combination at a given concentration, to represent the volume of the datasets in the following context. DrugCombDB includes dozens of large-scale wet-lab experimental datasets collected from publications and public resources. Most datasets come from drug combination screening projects funded by the National Institutes of Health (NIH), while others are collected from publications that focus on the discovery of combinatorial therapies.

A major part of the experimental data comes from the data portals run by the NIH. Thanks to the effort of the NCI-ALMANAC project (A Large Matrix of Anti-Neoplastic Agent Combinations) (18), we downloaded the large-scale drug combination dataset available on the National Cancer Institute (NCI) data portal. This dataset includes a  $3 \times 3$  dose-response matrix for each of 311 604 dual-agent combinations on 60 well-characterized human tumor cell lines. In total, there are 2873 514 experimental data points integrated into DrugCombDB. Additionally, the NCATS Matrix run by the NIH has deposited more than 10 drug combination screening datasets, some of which have been published. For example, the My-T-BCR supercomplex project carried out experiments to seek synergistic agents of ibrutinib (BTK inhibitor) from 30 mTOR inhibitors to treat diffuse large B-cell lymphoma (DLBCL). Their experiments output a  $10 \times 10$  dose-response matrix for each drug combination on the TMD cell line at two time points, resulting in 6000 combination tests. Additionally, the rhabdomyosarcoma project employed screening experiments to find small-molecule agents used in combination with trametinib to decrease rhabdomyosarcoma cell viability and slow tumor growth (19). They designed a  $10 \times 10$  dose-response block for each candidate of 96 combinations on three cell lines, thereby generating 28 800 combination tests. In addition, quite a few datasets that have not been published but have been submitted to the NCATS data portal are also integrated into DrugCombDB.

Experimental datasets collected from publications have also contributed substantially to our database. One large-scale dataset comes from an unbiased oncology compound screening experiment that was designed to identify combination strategies (20). This HTS was performed on a fully automated GNF PolyTarget robotic platform, where cancer cells were treated with a  $4 \times 4$  matrix of drug concentration combinations. Based on the cell viability measured using CellTiter-Glo cell viability reagent (Promega), the highest single agent (HSA) and Bliss independence models were applied to determine the combinatorial efficacy. In total, this assay yielded 1475 328 combination tests of 92 208 dual-drug combinations over 39 diverse cancer cell lines. Another HTS platform was adopted to uncover therapeutic combinations for the activated B-cell-like subtype (ABC) of DLBCL (21). This assay aimed to seek partner agents in collaboration with ibrutinib, a Bruton’s tyrosine kinase inhibitor, to treat ABC DLBCL. In total, 466 different agents were

evaluated in combination with ibrutinib using  $6 \times 6$  dose-response blocks, which outputs a total of 16 776 combination tests. Some relatively small datasets are also integrated into DrugCombDB. For example, Mohammad *et al.* explored the adaptive resistance of melanoma cells to RAF inhibitor (22) covering 25 different agents ( $5 \times 4$  dose-response block).

### AstraZeneca–Sanger drug combination dataset

The well-known pharmaceutical enterprise AstraZeneca partnered with the European Bioinformatics Institute, the Sanger Institute, Sage Bionetworks and the distributed DREAM community to launch the AstraZeneca–Sanger Drug Combination Prediction Challenge in 2015, to learn a set of general patterns or rules that could be used to predict synergistic behaviors in new compounds or disease contexts (23). As the data provider of the challenge, AstraZeneca released 11 576 experimentally tested drug combinations measuring cell viability over 118 drugs and 85 cancer cell lines. To extend the coverage of our database, we consulted with the data provider and obtained permission to integrate this large-scale dataset into DrugCombDB.

In summary, DrugCombDB contains 6055 926 dose combination tests covering 448 555 dual-drug combinations, 2887 unique drugs and 124 cancer cell lines. Detailed data statistics are outlined in Table 1.

### Manual literature curations

Many drug combinations under clinical and preclinical trials have been reported in the PubMed literature, and a large number of research papers have also investigated drug combinations with potential clinical effects evaluated by low-throughput biochemical assays, such as *in vitro* models and flow cytometry. To extend the coverage of our database, we work hard to manually review thousands of articles from PubMed to extract literature-supported drug combinations. Specifically, using ‘drug combination(s)’, ‘combination drug(s)’, ‘combinatorial drug(s)’ and ‘synergistic drug(s)’ as query keywords, we searched the PubMed database and obtained 8123 distinct publications that included at least one of these keywords in their titles or abstracts. Subsequently, we adopted PubTator (24), a web-based tool for facilitating manual literature curation through powerful text-mining techniques, to annotate the abstracts of these publications. Taking the PubMed ID list of the filtered publications as the input, PubTator marks discriminative conceptual keywords such as ‘gene’, ‘chemical’, ‘disease’, ‘species’ and ‘mutations’ in different colors. Subsequently, we manually reviewed the highlighted concepts and deeply investigated the context to identify therapeutically efficient drug combinations reported in these publications. Table 2 shows the number of retrieved PubMed publications corresponding to different keywords, as well as the number of final curated drug combinations, involved single agents and diseases. It is worth noting that these manually curated drug combinations from publications will be highly promising for developing combinatorial therapies for cancer treatment following approval for clinical use and therefore greatly enrich the value of our database.

### External databases

Several existing databases contain additional numbers of drug combinations, such as DCDB (16), DrugCentral (25), TTD (26), ASDCD (27) and DrugBank (28). DCDB 2.0 also collects 1363 drug combinations extracted from the FDA Orange Book (330 approved and 1033 investigational, including 237 unsuccessful usages), involving 904 individual drugs and 805 targets. DrugCentral latest version also includes 7621 combination sets covering 970 unique active ingredients with pharmaceutical formulations extracted from FDA Orange Book. The latest version of the TTD includes 72 pharmacodynamic synergistic combinations. ASDCD is a database of antifungal synergistic drug combinations, from which we collected 548 pairwise validated combinations. DrugBank also provides a dataset containing 13 397 clinically reported adverse drug combinations categorized as antagonistic. We removed duplicates among these external databases, and provided download links in our website.

### Multiple-agent combinations

Due to the limitations of HTS platforms and the combinatorial explosion of three or more component agents, most experiments consider only dual-agent combinations. However, complicated drug combinations mean more targets, which potentially improve efficacy and overcome resistance in the treatment of complex diseases and tumors (29). During the process of data collection, we also collected a few combinations composed of multiple agents reported by biochemical experiments or clinical trials. To expand the dimension of DrugCombDB, we also integrated these data into our database for potential usage.

## DATA PROCESSING AND INTEGRATION

### Identifier conversion

Because drug combinations are collected from various resources, the identifiers of drugs and cell lines vary across the different sources. For example, some experimental datasets use canonical drug names, while others use custom chemical identifiers. Therefore, the drug identifiers must be made uniform so that we can integrate these datasets. We choose the PubChem compound identifiers (CIDs) as uniform identifiers because they are widely used and easily linked to external public resources. Specifically, CIDs enable us to fetch detailed information, such as chemical structures, SMILES and pharmacological actions, from STITCH (30), PubChem (31) and DrugBank (28). We transferred the drug names to CIDs using the PubChem identifier exchange service. If one drug name corresponded to multiple CIDs, we chose the canonical name, i.e. the shortest CID code. Finally, for those names that could not be converted by the identifier exchange service, we manually retrieved them through search engines and other databases to ensure that these drugs were correctly converted to PubChem CIDs.

Similarly, the cell line identifiers are also inconsistent among different datasets. We mapped each cell line to COSMIC ID using COSMIC supplementary files (32), as well as Cellosaurus ID run by the ExPASy resource portal (33), so that we can associate cancer cell lines with external public



**Table 1.** Statistics of drug combinations collected from HTS assays

Dataset name	Landscape	Combination test	Cell line	Agent	Block size	PMID
Adult T-cell leukemia/lymphoma	466	16 776	1	464	6 × 6/10 × 10	
DLBCL	96	9600	1	17	10 × 10	
Diffuse intrinsic pontine gliomas	8703	450 908	2	2441	6 × 6/10 × 10	
Ebola	236	8496	2	17	6 × 6	29939303
Ewing's sarcoma	5952	219 904	2	1910	6 × 6/10 × 10	
GBM oncospheres	482	20 360	1	31	6 × 6/10 × 10	
Hodgkin's lymphoma	2648	110 688	4	1910	6 × 6/10 × 10	
MDR-CS	68	6800	2	19	10 × 10	
Malaria	13 325	669 716	3	224	6 × 6/10 × 10	
Rhabdomyosarcoma	288	28 800	3	27	10 × 10	29973406
CEPT	412	41 200	1	30	6 × 6/10 × 10	
ALMANAC (NCI)	311 604	2873 514	60	105	3 × 3/3 × 5	28446463
O'Neil	92 208	1475 328	39	38	4 × 4	26983881
Mathews	466	16 776	1	464	6 × 6	24469833
Mohammad	25	500	5	6	5 × 4	28069687
AstraZeneca DREAM challenge	11 576	106 560	85	118	6 × 6	31209238

**Table 2.** Statistics of drug combinations manually curated from the PubMed literature

Search keyword	Retrieved publication	Drug combination	Agent	Disease
Drug combination(s)	6531	179	688	242
Synergistic drug(s)	364	154		
Combinatorial drug(s)	166	43		
Combination drug(s)	1537	81		

resources. COSMIC ID allows users to obtain drug sensitivity profiles of cancer cell lines from the GDSC (34), while Cellosaurus IDs are linked to their corresponding diseases via EXPASY.

### Computation of synergy scores

It is well known that the efficacy of drug combinations may be synergistic or antagonistic, depending on whether cancer cells are inhibited or promoted to proliferation compared to the additive effect of independent treatments with single agents. If the percentage of inhibited or killed cancer cells is greater than expected, the drug combinations are classified as synergistic. On the other hand, antagonism is determined if more adverse effects are observed than expected. Mathematically, the combinatorial effect can be determined by the deviation of the dose–response curves from the expectation effect calculated based on a reference model. The Loewe additivity and Bliss independence models are the most frequently used of these reference models. Based on the raw dose–response landscape, we calculated multiple quantitative synergy scores based on different models, including the HSA, Loewe additivity, Bliss independence and zero interaction potency (ZIP).

For dual-drug combinations, the HSA model reflects the extent to which the resulting effect of a drug combination ( $E_{AB}$ ) is greater than the maximum effect produced by individual component drugs A ( $E_A$ ) and B ( $E_B$ ) (35), and its combinatorial index corresponding to one combination test can be calculated as follows:

$$E_{HSA} = \frac{E_{AB}}{\max(E_A, E_B)}. \quad (1)$$

Note that Equation (1) actually computes the combinatorial index corresponding to one combination test of a single dose–response landscape. The final HSA synergy score

is defined as the mean combination index over all combinatorial dose responses except for those treated by a single agent.

The Bliss independence model (36) assumes that drugs act independently so that neither interferes with each other but each contributes to the final effect. The combination index of the Bliss model is expressed as the ratio of the observed combination effect to the expected additive effect:

$$E_{Bliss} = \frac{E_A + E_B - E_A \times E_B}{E_{AB}}. \quad (2)$$

The Loewe additivity model also assumes that the effects of two drugs are independent. However, it also relies on both the dose-equivalence principle and the mimic combination principle (the individual drug dose responses must be monotonic). Furthermore, the Loewe additivity also allows us to complement the algebraic analysis with an intuitive, flexible and widely accepted graphical approach known as isobologram analysis (37). The Loewe combination index is defined as follows:

$$E_{Loewe} = \frac{E_{\min} + E_{\max}[(d_A + d_B)/m]^\lambda}{1 + [(d_A + d_B)/m]^\lambda}, \quad (3)$$

where  $d_A$  and  $d_B$  are the doses of drug A and drug B, respectively,  $E_{\min}$  and  $E_{\max}$  are the minimal and maximal effects of the drug combination ( $0 \leq E_{\min} < E_{\max} \leq 1$ ), respectively,  $m$  is the dose that produces the midpoint effect of  $E_{\min} + E_{\max}$ , i.e. the relative  $EC_{50}$  or  $IC_{50}$ , and  $\lambda$  ( $\lambda > 0$ ) is the slope of the curve.

In addition to the three classical reference models mentioned earlier, we also adopted a novel reference model named ZIP to compute another synergy score. The ZIP model has been demonstrated to capture the drug interaction relationships by comparing the changes in the potency of the dose–response curves between individual drugs and

their combinations (38). ZIP is a response surface model that combines the advantages of the Loewe and the Bliss models, which proposes a delta score to characterize the synergy landscape over the full dose–response matrix. The ZIP model assumes that two noninteracting drugs are expected to incur minimal changes in their dose–response curves. A delta score is computed to quantify the deviation from the expectation of ZIP for a given dose pair and utilizes the average delta score over a dose–response matrix as a summary interaction score for a particular drug combination. As a result, the ZIP model is perfectly compatible with high-throughput drug combination screening data. The ZIP combination index is defined as follows:

$$E_{ZIP} = \frac{(d_A/m_A)^{\lambda_A}}{1 + (d_A/m_A)^{\lambda_A}} + \frac{(d_B/m_B)^{\lambda_B}}{1 + (d_B/m_B)^{\lambda_B}} - \frac{(d_A/m_A)^{\lambda_A}}{1 + (d_A/m_A)^{\lambda_A}} \times \frac{(d_B/m_B)^{\lambda_B}}{1 + (d_B/m_B)^{\lambda_B}}, \quad (4)$$

where  $d_A$  and  $d_B$  are the doses of drug A and drug B, respectively,  $m_A$  and  $m_B$ , respectively, are the doses that produce the midpoint effect when using drug A at  $d_A$  or drug B at  $d_B$  individually, and  $\lambda_A$  and  $\lambda_B$  are the slopes of the curves induced by the individual drugs, respectively.

Note that the combination index computed by each reference model accounts for only one combination test. The final synergy score is actually the mean combination index over all dose responses except for the single-agent treatment ones. In our implementation, we applied the R package SynergyFinder (39) to calculate the quantitative synergy scores of the four reference models.

### Normalization of dose responses

Due to the heterogeneity of the different platforms on which the HTS assays were performed, the magnitudes of the dose response levels vary across the different experimental protocols and techniques. To facilitate downstream usage of our database, we normalized the dose response levels to provide coincident and comparable therapeutic efficacy over different datasets. Considering that cell viability and apoptosis rate upon treatment are the most commonly used measures in drug sensitivity assays, we computed the inhibition rate ( $R_{inhibit}$ ) of cancer cells to drug treatments as a uniform measure using the min–max normalization method:

$$R_{inhibit} = \frac{\max(\text{viability}) - \text{viability}}{\max(\text{viability}) - \min(\text{viability})}, \quad (5)$$

$$R_{inhibit} = \frac{\text{apoptosis} - \min(\text{apoptosis})}{\max(\text{apoptosis}) - \min(\text{apoptosis})}. \quad (6)$$

Min–max normalization was applied to each separate dataset before integration. As a result, 1 represents the highest sensitivity and 0 represents the lowest sensitivity. Because the normalized inhibition rates range from 0 to 1, this is very favorable for downstream usage. In fact, the normalized inhibition rates over different datasets greatly expand the volume of data for modeling techniques such as regression analysis, especially taking different drug concentration combinations into account.

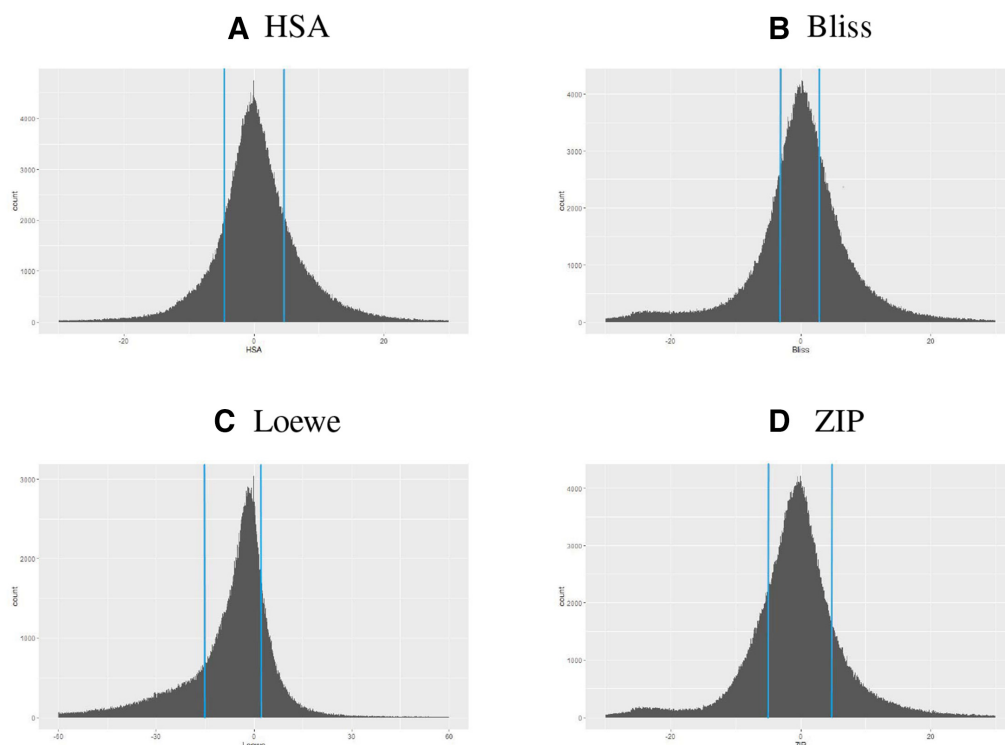
### Classification of synergism and antagonism

Theoretically, drug combinations can be roughly classified as synergistic and antagonistic according to the aforementioned synergy scores, i.e. the combination of drugs is more or less effective in killing cancer cells than addition of efficacy induced by independent administration of individual drugs. Synergism and antagonism can be easily decided with a cutoff of zero. The higher the scores, the more synergistic the combination, and vice versa. However, as shown in Figure 1, the synergy scores derived from the four synergy scoring models follow nearly normal distributions, with most samples located close to zero, indicating that a zero cutoff cannot clearly discriminate the combinatorial effect. Furthermore, the synergy scoring models are not sufficiently robust to the noise in HTS experiments; therefore, drug combinations with synergy scores located near zero are difficult to classify. To tackle this dilemma, the z-score is often computed, and then relatively strict thresholds are used to exclude low-confidence samples. Here, we adopted the quartile as the threshold; namely, approximately one-quarter of the combinations with higher scores and one-quarter of the combinations with lower scores were classified as synergistic and antagonistic. For the Loewe synergy scores, we notice that its distribution is significantly biased to the negative side. To balance the number of synergistic and antagonistic samples, we adjust its threshold for antagonism to a relatively lower level. The thresholds used in classifying each type synergy score are highlighted by blue vertical bars, as shown in Figure 1.

Furthermore, we took into account all four synergy scores together and classified the drug combinations based on a majority voting strategy. For a particular drug combination, if and only if each of the four synergy scores indicated synergism (antagonism), it was labeled as synergistic (antagonistic). As a result, 85 154 synergistic and 155 824 antagonistic drug combinations were filtered out. We hope that these prebuilt datasets will facilitate bioinformaticians to develop innovative *in silico* methods for the prediction of drug combinations. Note that the prebuilt datasets mentioned earlier are available on the download page of our website.

### Integration of drug combination replicates

There are a number of replicates of drug combinations within one dataset and among different datasets. It is necessary to check the reproducibility of replicates before the integration of these replicates of drug combinations. Specifically, we introduced drug combination sensitivity score (CSS) (40), which is designed to integrate the sensitivity and dose–response synergy of combination tests, to evaluate the reproducibility between the replicates within a dataset. Let us take the O’Neil and NCI-ALMANAC for example, as they are the top 2 largest datasets among the experimental data sources. The standard deviation (SD) probability density curves of CSS scores of the replicates within O’Neil and ALMANAC and between these two datasets are shown in Figure 2. Note that the average SD for O’Neil (22 737 combinations are replicated for one or more times) is 3.44, which is significantly lower than other two SD values, suggesting a satisfactory reproducibility for the replicates within



**Figure 1.** Distribution of the synergy scores computed using four reference models. Parts (A)–(D) correspond to HSA, Bliss, Loewe and ZIP, respectively. The blue vertical bars are the quartile thresholds for the classifications used for each type of synergy score.

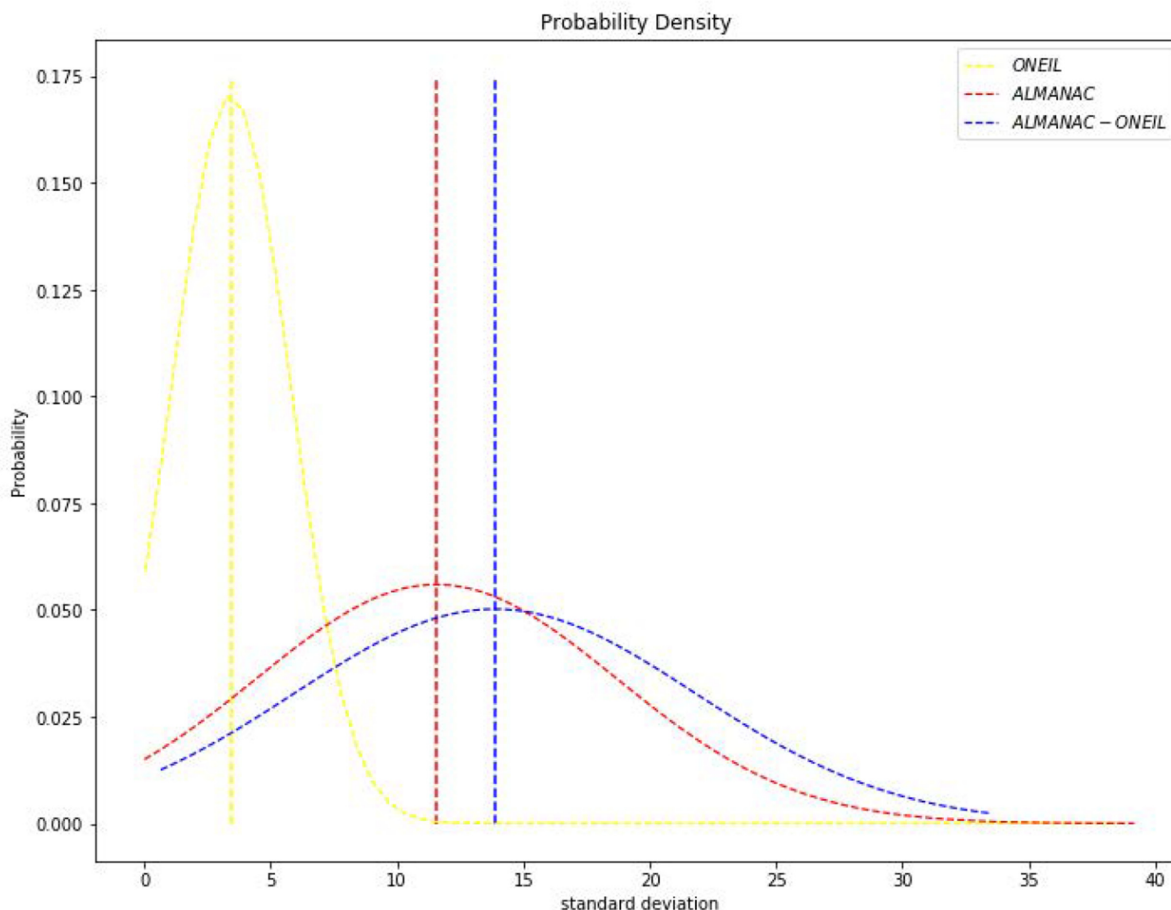
the O’Neil dataset. In addition, we find that all the replicates were assayed under same dosages within the O’Neil dataset; therefore, each replicate can be regarded as the evaluation assay with same full dose–response matrix. To obtain a unique synergy score for each drug combination, we calculated the average inhibition levels for same drug pair on same type of cancer cell line to get the dose–response matrix.

The average SD of replicate within ALMANAC dataset (2091 combinations have replicates) is 11.56, and the average SD of the between-study replicates (352 combinations) is 13.867, which are both higher than that of O’Neil dataset. We subsequently went through the raw dose response levels of these combinations, and found that a majority of the assayed doses of these replicates are different from each other, which may result in the relatively unstable synergy scores. Note that the average SD values within ALMANAC and between-study replicates are still significantly lower than the SD (24.76) of CSS of the union set of O’Neil and ALMANAC, suggesting that the CSS scores of replicates are significantly closer than random samples. As a result, we still believe that the reproducibility is acceptable. Considering the small quantity of these combinations in ALMANAC and between studies, we do not combine the synergy score of these replicates, keeping them for further assessment.

## FUNCTIONALITIES

A website with a user-friendly interface is provided to make full advantage of the wealth of data. We have devel-

oped a few functional modules, including ‘search’, ‘filtering’, ‘graphical visualization’ and ‘download’ modules. The search module takes either a single drug or drug combination of interest as input to search the drug combinations, and the search results are presented in a tabular viewer, as shown in Figure 3. Each row of the table lists a drug combination, the individual agents, the cell line, multiple synergy scores and the data sources. Note that the drug combination is represented by separating the individual drug names using a minus sign. The tabular viewer functions as a portal to explore more detailed information related to drug combinations. First, to help understand the pharmacological actions of the drug combination, we display their common target proteins that are collected from various data source of drug–target interactions. DrugBank (28) provides targets of thousands of FDA-approved and experimental drugs. STITCH (30) gives comprehensive drug–target networks that combine multiple supporting evidences and compute a confidence score for each drug–target interaction. In particular, recent studies have built the mapping of oncology drugs and their efficacy targets (41). For example, Santos *et al.*’s review paper released 4631 drug–target relationships between 1578 unique drugs and 667 human biomolecules (41), and each relationship has been annotated with mechanism of action, protein homologous families and modulation. Lin *et al.* adopted CRISPR–Cas9 mutagenesis to investigate whether a set of cancer drugs really target their putative targets or not (42). We integrate all these drug–target relationships so that we can search common targets of individual drugs in a combination set. Specifically, users can click the drug combination item to browse the top 10 tar-



**Figure 2.** The probability density curve plot of replicated combinations.

get proteins in common between the individual drugs in a graphical network viewer, in which rectangles represent the drugs and circles represent their common target proteins. The table located on the right-hand side shows the properties of the component drugs, including molecular weight, chemical structure, SMILES and external links to STITCH and PubChem. The full list of common target proteins of the drug combinations, together with the confidence scores derived from STITCH, is shown in Figure 4. Also, we highlight the drug targets using differently colored labels corresponding to different data sources supporting the drug–target relationships. Users can also click on the individual drugs to show their detailed information and full list of targets in the form of both graphical network graphs and tabular viewers. Particularly, to encourage further research, we have exploited DrugCentral (25) to annotate the drugs in our database, and the label of ‘approved’ or ‘investigational’ will be displayed beside the drug name. The cell line names link to pages that show more details about the cell line and associated disease. In particular, the sensitivities of the cell line to the individual drugs are collected from GDSC and shown in a scatter plot, in which the sizes of the circles are proportional to the sensitivity measures ( $IC_{50}$  values), as shown in Figure 5. Mousing over a circle will show details about the cell line and sensitivity. It is worth noting that this

function enables the user to conduct a comparison of therapeutic efficacy between the drug combination and single-component drugs.

Moreover, users can click on a synergy score to explore the raw dose–response landscape. The raw dose responses are collected from different HTS assays, and the block sizes are dependent on the HTS platform used and the experimental design. For instance, Figure 6 shows the heatmap of the raw dose–response matrix of the combination of chlorambucil and thioguanine on the cell line ACHN. Each cell of the heatmap represents the viability (e.g.  $IC_{50}$ ) of the cancer cells when treated by the drug combination at the corresponding concentrations of the single agents, and its color depth corresponds to the sensitivity value. Meanwhile, the  $\Delta$  Bliss matrix shows the score matrix computed using the Bliss independence model.

The ‘filtering’ module helps users rapidly find drug combinations regarding diseases of interest. With the filter panel, user can choose one or more tissues to highlight associated cancer cell lines and then check the cell lines to filter drug combinations within the search results. To facilitate rapid input of drug names, an ‘autocomplete’ function was implemented to provide contextual completion functionality for custom text matches based on user input. Every time a user types a preconfigured special character, he



Drug Combinations

Input drug name(s):

Filter by tissue:

Brain
  Lung
  Liver
  Misc
  Haematopoietic\_and\_Lymphoid
  Kidney
  Ovary
  Uterus  
 Rectum
  Breast
  Skin
  Colon
  Bone
  Prostate

Id	DrugCombination	Drug1	Drug2	CellName	HSA	Bliss	Loewe	ZIP	Source
82944	MK-4827 - Metformin	Metformin	MK-4827	A2058	1.3425	2.9325	-27.575	6.095	pmid26983881
83200	ABT-888 - Temozolomide	Temozolomide	ABT-888	A2058	16.2225	16.4925	11.9925	16.2625	pmid26983881
83456	Temozolomide - L778123	L778123	Temozolomide	A2780	7.5425	6.405	4.635	5.9975	pmid26983881
83712	Topotecan - MRK-003	MRK-003	Topotecan	A2780	13.8725	10.3225	10.1025	11.8475	pmid26983881
83968	MK-8776 - Dexamethasone	Dexamethasone	MK-8776	A375	9.2975	12.7375	-6.1425	9.05	pmid26983881
84224	SN-38 - MK-4827	MK-4827	SN-38	A375	19.83	15.3825	15.2225	15.7825	pmid26983881
84480	MK-2206 - Carboplatin	Carboplatin	MK-2206	A427	5.995	-0.4225	3.6775	-1.435	pmid26983881
84736	Topotecan - AZD1775	AZD1775	Topotecan	A427	10.3925	4	4.5725	4.7925	pmid26983881
84992	geldanamycin - Zolanza	Zolanza	geldanamycin	A427	-6.055	-11.5875	-9.6025	-7.5925	pmid26983881
85248	Lapatinib - Metformin	Metformin	Lapatinib	CAOV3	3.0525	2.785	-26.74	-3.8075	pmid26983881

Figure 3. Screenshot of a tabular view of drug combinations, as well as the search and filtering panels.

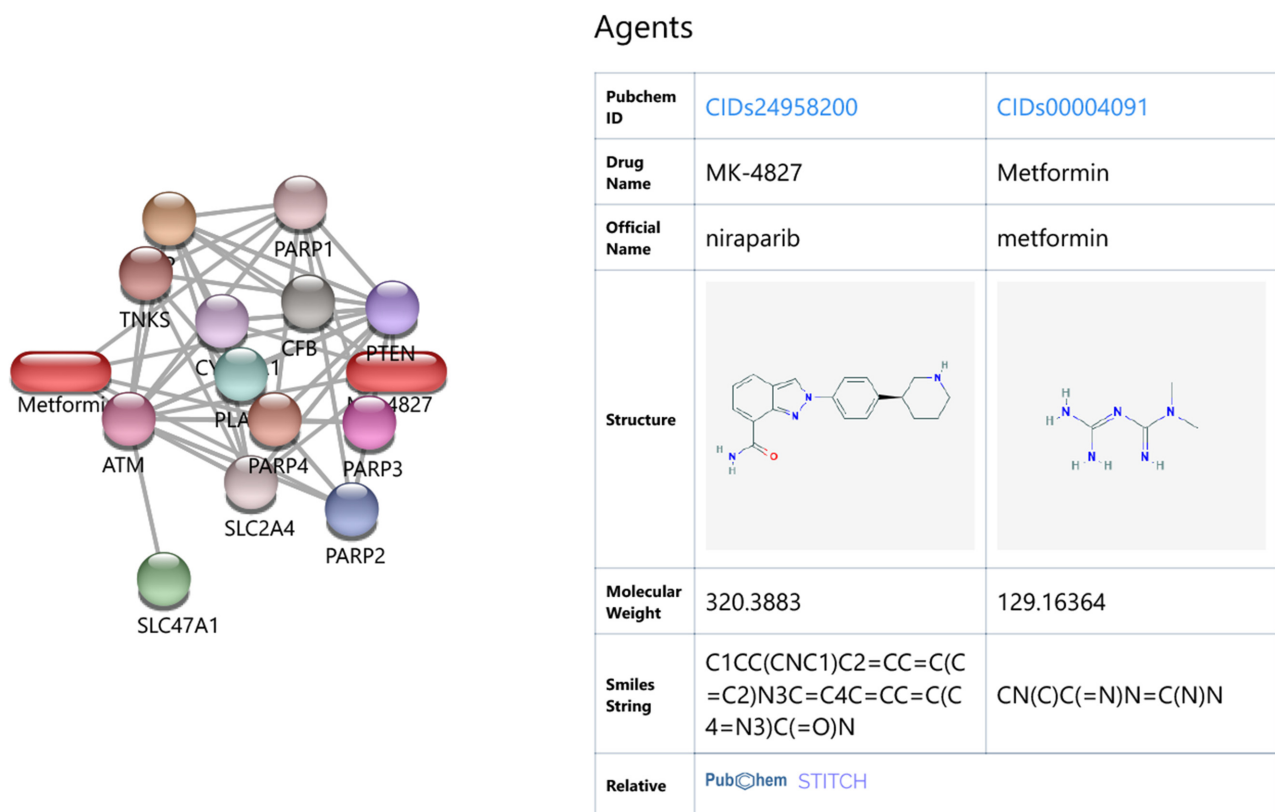
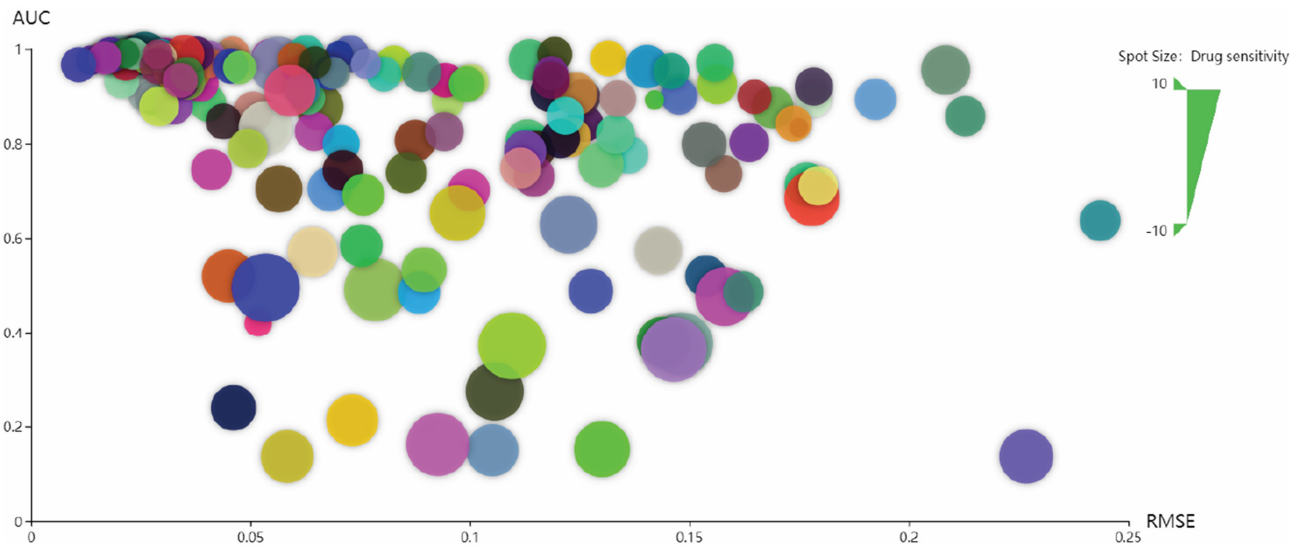
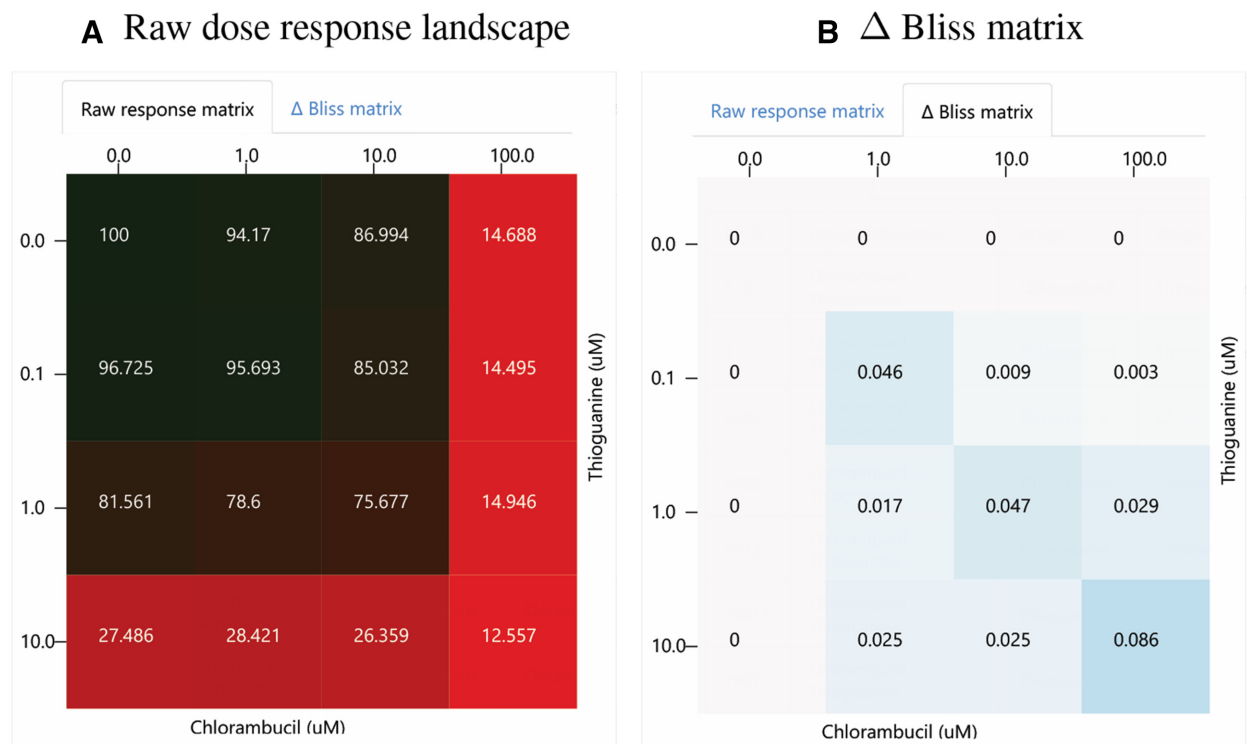


Figure 4. Common targets of the drug combination MK-4827 and metformin in a graphical network viewer. The table on the right-hand side shows the chemical properties of the single agents.





**Figure 5.** Screenshot of the scatter plot of the drug sensitivity of the cancer cell line to the individual drugs. Note that the drug sensitivity data come from the GDSC and that the sizes of the circles correspond to the inverse of the log IC<sub>50</sub> values.



**Figure 6.** Heatmap of the raw dose–response landscape (A) and  $\Delta$  Bliss matrix (B) of the combination of drugs chlorambucil and thioguanine on cell line ACHN.

or she will obtain indications about available autocomplete suggestions displayed in a dedicated drop-down window. In addition, with the ‘ranking’ function of the tabular viewer, users can easily select drug combinations with high sensitivity toward the cancer cells of interest with respect to certain types of synergy scores, such as Bliss, Loewe or ZIP.

DrugCombDB also aims to help the academic and industrial community obtain our datasets for further analy-

sis, such as wet-lab screening experiments and *in silico* prediction algorithms. Therefore, we developed a ‘download’ page through which the user can obtain any dataset of interest. For example, all drug combinations with quantitative synergy scores can be downloaded and classified using custom thresholds into synergism and antagonism. Our prebuilt sets of drug combinations already categorized into synergism and antagonism according to synergy scores can

also be obtained. To facilitate the construction of a gold standard set of drug combinations, all FDA-approved drug combinations have been released. Moreover, the literature-supported and clinically tested drug combinations are also available. We expect that the *in silico* prediction of synergistic drug combinations or preclinical experiments may benefit from our datasets.

## DISCUSSION AND CONCLUSION

With the exponentially increasing amount of pharmacology and transcriptomic data, there is a pressing need to create *in silico* methods to predict synergistic drug combinations that can satisfy personalized cancer treatments and combat increased drug resistance. However, the current drug combination databases are limited in function when applied in novel algorithms because of their small sizes. Therefore, the DrugCombDB presented here can extensively facilitate the training and validation of many advanced machine learning algorithms by heuristically providing various types of datasets.

Our database contains a large amount of data on experimental and well-documented drug combinations. We think these data would greatly facilitate both wet-lab and *in silico* researchers. First, the computed synergy score would provide effective indications on whether the drug combinations have the potential to be developed into a viable clinical therapy. We believe that our database will function to narrow down the number of drug combinations for further experimental study. However, it is worth noting that the synergy scores only reflect the overall synergistic or antagonistic effect based on dose–response landscapes, leading to the fluctuation by the changes of assayed dosage concentrations. According to Figure 1, the distributions of all four types of synergy scores follow nearly normal distributions with most samples located close to zero, indicating that a zero cutoff cannot clearly discriminate the combinatorial effect. As a result, the weight of synergy scores in the classification of synergistic or antagonistic is limited.

Second, we have constructed a few datasets of classification of synergism and antagonism according to canonical thresholds to facilitate subsequent usage of the datasets. More importantly, we have collected a large number of extra combination sets from FDA Orange Book, clinicaltrials.gov database and PubMed literature followed by manual curations. This set of drug combinations includes high-confidence ones compared to HTS experimental results, so that we can build a gold standard set of drug combinations that is particularly important in the development of computational methods that highly depends on the quality of training data.

Last but not least, we have received much attention since the release of our database and bioRxiv preprint in December 2018. We believe that the database will greatly facilitate both the academic community and industrial enterprises. Future efforts include collecting more characteristic information on cancer cell lines and pathways, such that researchers can tap into the biological mechanisms more deeply. Users are welcome to participate in the extension of the project and help DrugCombDB become more com-

prehensive, which may contribute to the development of innovative combinatorial therapies.

## ACKNOWLEDGEMENTS

We would like to thank Dr Michael Mason for his kind approval of our application for integration of the AstraZeneca–Sanger drug combination dataset into our database. Note that the AstraZeneca–Sanger drug combination dataset can only be browsed but not bulk downloaded from the website.

## FUNDING

National Natural Science Foundation of China [61672113, 61972422, 61672541]. Funding for open access charge: National Natural Science Foundation of China [61672113, 61972422, 61672541].

*Conflict of interest statement.* None declared.

## REFERENCES

- Caitriona,H., Sandra,V., Daniel,B.L. and Patrick,G.J. (2013) Cancer drug resistance: an evolving paradigm. *Nat. Rev. Cancer*, **13**, 714–726.
- Michael,D., Tito,F. and Susan,B. (2005) Tumour stem cells and drug resistance. *Nat. Rev. Cancer*, **5**, 275–284.
- Pang,K., Wan,Y., William,T.C., Lawrence,A.D., Sun,J., Dhruv,P. and Liu,Z. (2014) Combinatorial therapy discovery using mixed integer linear programming. *Bioinformatics*, **30**, 1456–1463.
- Michael,M.G., Orit,L., Matthew,D.H. and Jean-Pierre,G. (2016) Toward a better understanding of the complexity of cancer drug resistance. *Annu. Rev. Pharmacol. Toxicol.*, **56**, 85–102.
- Cheng,F., István,A.K. and Albert-László,B. (2019) Network-based prediction of drug combinations. *Nat. Commun.*, **10**, 1197.
- Jia,J., Zhu,F., Ma,X., Cao,Z., Cao,Z., Li,Y., Li,Y. and Chen,Y. (2009) Mechanisms of drug combinations: interaction and network perspectives. *Nat. Rev. Drug Discov.*, **8**, 111–128.
- Bissan,A., Udai,B. and Paul,W. (2012) Combinatorial drug therapy for cancer in the post-genomic era. *Nat. Biotechnol.*, **30**, 679–692.
- Péter,C., Vilmos,A. and Sandor,P. (2005) The efficiency of multi-target drugs: the network approach might help drug design. *Trends Pharmacol. Sci.*, **26**, 178–182.
- Sergio,I., Lu,Y., Fabiana,C.M., Gordon,B.M. and Prahlad,T.R. (2010) Identification of optimal drug combinations targeting cellular networks: integrating phospho-proteomics and computational network analysis. *Cancer Res.*, **70**, 6704–6714.
- Tang,J., Leena,K., Xu,T., Agnieszka,S., Bhagwan,Y., Krister,W. and Tero,A. (2013) Target inhibition networks: predicting selective combinations of druggable targets to block cancer survival pathways. *PLoS Comput. Biol.*, **9**, e1003226.
- Michael,J.L., Ye,S.A., Alexandra,K.G., Anne Margriet,H., Peter,K.S., Gavin,M. and Michael,B.Y. (2012) Sequential application of anticancer drugs enhances cell death by rewiring apoptotic signaling networks. *Cell*, **149**, 780–794.
- Guo,J., Liu,H. and Zheng,J. (2015) SynLethDB: synthetic lethality database toward discovery of selective and sensitive anticancer drug targets. *Nucleic Acids Res.*, **44**, D1011–D1017.
- James,C.C., Laura,M.H., Elisabeth,G., Mehmet,G., Michael,P.M., Nicholas,J.W., Mukesh,B., Petteri,H., Suleiman,A.K., John-Patrick,M. *et al.* (2014) A community effort to assess and improve drug sensitivity prediction algorithms. *Nat. Biotechnol.*, **32**, 1202–1212.
- Sun,X., Santiago,V. and Nicholas,P.T. (2013) High-throughput methods for combinatorial drug discovery. *Sci. Transl. Med.*, **5**, 205rv1.
- Bilal,M., Wang,W., Wang,J., Howard,C., Neo Christopher,C. and Ping,P. (2019) Machine learning and integrative analysis of biomedical big data. *Genes*, **10**, e87.
- Liu,Y., Wei,Q., Yu,G., Gai,W., Li,Y. and Chen,X. (2014) DCDB2.0: a major update of the drug combination database. *Database*, **2014**, bau124.

17. Bulat,Z., Jehad,A., Zheng,S., Wang,W., Wang,Y., Joseph,S., Alina,M., Mohieddin,J., Ziaurrehman,T., Alberto,P. *et al.* (2019) DrugComb: an integrative cancer drug combination data portal. *Nucleic Acids Res.*, **47**, W43–W51.
18. Susan,L.H., Richard,C., James,A.C., Jeevan Prasaad,G., Melinda,H., Lawrence,W.A., Eric,P., Larry,R., Apurva,S., Deborah,W. *et al.* (2017) The National Cancer Institute ALMANAC: a comprehensive screening resource for the detection of anticancer drug pairs with enhanced therapeutic activity. *Cancer Res.*, **77**, 3564–3576.
19. Marielle,E.Y., Berkley,E.G., Jack,F.S., Song,Y.K., Hsien-Chao,C., Sivasish,S., Arnulfo,M., Rajesh,P., Zhang,X., Rajarashi,G. *et al.* (2018) MEK inhibition induces MYOG and remodels super-enhancers in RAS-driven rhabdomyosarcoma. *Sci. Transl. Med.*, **10**, eaan4470.
20. O’Neil,J., Benita,Y., Feldman,I., Chenard,M., Roberts,B., Liu,Y., Li,J., Kral,A., Lejnine,S., Loboda,A. *et al.* (2016) An unbiased oncology compound screen to identify novel combination strategies. *Mol. Cancer Ther.*, **15**, 1155–1162.
21. Mathews Griner,L.A., Guha,R., Shinn,P., Young,R.M., Keller,J.M., Liu,D., Goldlust,I.S., Yasgar,A., McKnight,C., Boxer,M.B. *et al.* (2014) High-throughput combinatorial screening identifies drugs that cooperate with ibrutinib to kill activated B-cell-like diffuse large B-cell lymphoma cells. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 2349–2354.
22. Mohammad,F., Verena,B., Benjamin,I., Gregory,J.B., Jia-Ren,L., Sarah,A.B., Parin,S., Asaf,R., Levi,A.G. and Peter,K.S. (2017) Adaptive resistance of melanoma cells to RAF inhibition via reversible induction of a slowly dividing de-differentiated state. *Mol. Syst. Biol.*, **13**, 905.
23. Michael,P.M., Dennis,W., Mike,J.M., Bence,S., Krishna,C.B., Guan,Y., Thomas,Y., Jaewoo,K., Minji,J., Russ,W. *et al.* (2019) Community assessment to advance computational prediction of cancer drug combinations in a pharmacogenomic screen. *Nat. Commun.*, **10**, 2674.
24. Wei,C., Kao,H. and Lu,Z. (2013) PubTator: a web-based text mining tool for assisting biocuration. *Nucleic Acids Res.*, **41**, W518–W522.
25. Oleg,U., Jayme,H., Cristian,G.B., Jeremy,J.Y., Stephen,L.M., Vasileios,S., Dac-Trung,N., Stephan,S. and Tudor,O. (2018) DrugCentral 2018: an update. *Nucleic Acids Res.*, **47**, D963–D970.
26. Li,Y., Yu,C., Li,X., Zhang,P., Tang,J., Yang,Q., Fu,T., Zhang,X., Cui,X., Tu,G. *et al.* (2017) Therapeutic target database update 2018: enriched resource for facilitating bench-to-clinic research of targeted therapeutics. *Nucleic Acids Res.*, **46**, D1121–D1127.
27. Chen,X., Ren,B., Chen,M., Liu,M., Ren,W., Wang,Q., Zhang,L. and Yan,G. (2014) ASDCD: antifungal synergistic drug combination database. *PLoS One*, **9**, e86499.
28. David,S.W., Yannick,D.F., Guo,A., Lo,E., Ana,M., Jason,R.G., Tanvir,S., Daniel,J., Li,C., Zinat,S. *et al.* (2017) DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.*, **46**, D1074–D1082.
29. Minoru,K., Yoko,S., Miho,F., Kanae,M. and Mao,T. (2018) New approach for understanding genome variations in KEGG. *Nucleic Acids Res.*, **47**, D590–D595.
30. Damian,S., Alberto,S., Christian,V., Lars Juhl,J., Peer,B. and Michael,K. (2015) STITCH 5: augmenting protein–chemical interaction networks with tissue and affinity data. *Nucleic Acids Res.*, **44**, D380–D384.
31. Sunghwan,K., Paul,A.T., Evan,E.B., Jie,C., Gang,F., Asta,G., Han,L., Jane,H., He,S., Benjamin,A.S. *et al.* (2015) PubChem substance and compound databases. *Nucleic Acids Res.*, **44**, D1202–D1213.
32. Simon,A.F., Nidhi,B., Sally,B., Charlotte,C., Chai Yin,K., David,B., Jia,M., Rebecca,S., Kenric,L., Andrew,M. *et al.* (2010) COSMIC: mining complete cancer genomes in the catalogue of somatic mutations in cancer. *Nucleic Acids Res.*, **39**, D945–D950.
33. Elisabeth,G., Alexandre,G., Christine,H., Ivan,I., Ron,D.A. and Amos,B. (2003) ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.*, **31**, 3784–3788.
34. Yang,W., Jorge,S., Patricia,G., Elena,J.E., Howard,L., Simon,F., Nidhi,B., Dave,B., James,A.S., I Richard,T. *et al.* (2012) Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.*, **41**, D955–D961.
35. Morris,C.B. (1989) What is synergy? *Pharmacol. Rev.*, **41**, 93–141.
36. CI,B. (1939) The toxicity of poisons applied jointly. *Ann. Appl. Biol.*, **26**, 585–615.
37. William,R.G., Gregory,B. and John,C.P. (1995) The search for synergy: a critical review from a response surface perspective. *Pharmacol. Rev.*, **47**, 331–385.
38. Bhagwan,Y., Krister,W., Tero,A. and Tang,J. (2015) Searching for drug synergy in complex dose–response landscapes using an interaction potency model. *Comput. Struct. Biotechnol. J.*, **13**, 504–513.
39. Aleksandr,I., He,L., Tero,A. and Tang,J. (2017) SynergyFinder: a web application for analyzing drug combination dose–response matrix data. *Bioinformatics*, **33**, 2413–2415.
40. Alina,M., Muntasir Mamun,M., Wang,W., Alberto,P., Caroline,A.H. and Tang,J. (2019) Drug combination sensitivity scoring facilitates the discovery of synergistic and efficacious drug combinations in cancer. *PLoS Comput. Biol.*, **15**, e1006752.
41. Santos,R., Ursu,O., Gaulton,A., Bento,A.P., Donadi,R.S., Bologa,C.G., Karlsson,A., Al-Lazikani,B., Hersey,A., Oprea,T.I. *et al.* (2017) A comprehensive map of molecular drug targets. *Nat. Rev. Drug Discov.*, **16**, 19–34.
42. Lin,A., CJ,G., Palladino,A., John,K., Abramowicz,C., Yuan,M., Sausville,E., Lukow,D., Liu,L., Chait,A. *et al.* (2019) Off-target toxicity is a common mechanism of action of cancer drugs undergoing clinical trials. *Sci. Transl. Med.*, **11**, eaaw8412.