



Computational frameworks transform antagonism to synergy in optimizing combination therapies



Jinghong Chen^{1,2,10}, Anqi Lin^{1,10}, Aimin Jiang^{3,10}, Chang Qi^{4,10}, Zaoqu Liu⁵, Quan Cheng^{6,7}✉, Shuofeng Yuan^{8,9}✉ & Peng Luo^{1,9}✉

While drug combinations are increasingly important in disease treatment, predicting their therapeutic interactions remains challenging. This review systematically analyzes computational methods for predicting drug combination effects through multi-omics data integration. We comprehensively assess key algorithms including DrugComboRanker and AuDNNsynergy, and evaluate integration approaches encompassing kernel regression and graph networks. The review elucidates artificial intelligence applications in predicting drug synergistic and antagonistic effects.

The synergistic and antagonistic interactions between drugs are fundamental concepts in combination drug therapy. When drugs are used in combination, if their combined effect exceeds the sum of the effects of each drug used individually, it is referred to as synergy¹. Conversely, if the effect of the drug combination is inferior to the sum of the individual drug effects, it is termed antagonism². Identifying synergistic drug combinations can significantly enhance therapeutic efficacy, while antagonistic combinations may lead to adverse consequences. Drug combination therapy has emerged as a promising approach to enhancing treatment outcomes, mitigating toxicity, and surmounting drug resistance, particularly when treating multifaceted diseases such as cancer. However, determining the optimal synergistic drug combinations remains a daunting challenge. Traditional experimental screening methods are laborious and resource-intensive, making it challenging to navigate the ever-expanding combinatorial space. Therefore, there is an imperative need to develop computational methods to predict drug synergy and guide the optimization of drug combinations.

In recent years, the rapid advancement of multi-omics technologies has offered unparalleled opportunities for systematically elucidating biological processes. The evolution of these technologies began with genomics in the 1990s through the Human Genome Project³, followed by transcriptomics in the early 2000s with microarray technology⁴, and later expanded to include proteomics, metabolomics, and epigenomics. This technological progression has created unprecedented opportunities for comprehensive biological

analysis. At present, numerous computational methods based on multi-omics data have been employed in predicting drug combinations. Through the integration of multi-level omics data (e.g., genomics, epigenomics, transcriptomics), these approaches can precisely predict drug interactions. The adoption of omics-based approaches was catalyzed by the Dialog on Reverse Engineering Assessment and Methods (DREAM) Drug Sensitivity Prediction Challenge⁵, which demonstrated the superior predictive power of genomic features. For example, Preuer et al. developed the DeepSynergy model, which incorporates compound chemical structures, gene expression profiles, and cell line information to predict drug synergies⁶. The model achieved impressive performance with a mean Pearson correlation coefficient of 0.73 between measured and predicted values, and an Area Under the Curve (AUC) of 0.90 for classification tasks, demonstrating a 7.2% improvement in mean squared error compared to other state-of-the-art methods⁶. This was followed by further innovations such as AuDNNsynergy in 2020 and more recent advances like PRODeepSyn (2022) and DGSSynADR (2023), showing the field's rapid evolution toward more sophisticated deep learning approaches. Huang et al. leverages gene expression profile data and protein-protein interaction (PPI) data to predict synergistic drug combinations. Huang et al. leverages gene expression profile data and PPI data to predict synergistic drug combinations⁷. These studies illustrate that computational modeling and prediction methods based on multi-omics data exhibit immense potential in drug combination

¹Department of Oncology, Zhujiang Hospital, Southern Medical University, Guangzhou, Guangdong, China. ²The Second School of Clinical Medicine, Southern Medical University, Guangzhou, Guangdong, China. ³Department of Urology, Changhai Hospital, Naval Medical University (Second Military Medical University), Shanghai, China. ⁴Vienna University of Technology, Institute of Logic and Computation, Vienna, Austria. ⁵Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China. ⁶Department of Neurosurgery, Xiangya Hospital Central South University, Changsha, Hunan, China. ⁷National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University, Changsha, Hunan, China. ⁸Department of Infectious Disease and Microbiology, The University of Hong Kong-Shenzhen Hospital, Shenzhen, Guangdong, China. ⁹Department of Microbiology, State Key Laboratory of Emerging Infectious Diseases, Carol Yu Centre for Infection, School of Clinical Medicine, Li Ka Shing Faculty of Medicine The University of Hong Kong, Hong Kong, China. ¹⁰These authors contributed equally: Jinghong Chen, Anqi Lin, Aimin Jiang, Chang Qi. ✉e-mail: chengquan@csu.edu.cn; yuanf@hku.hk; luopeng@smu.edu.cn

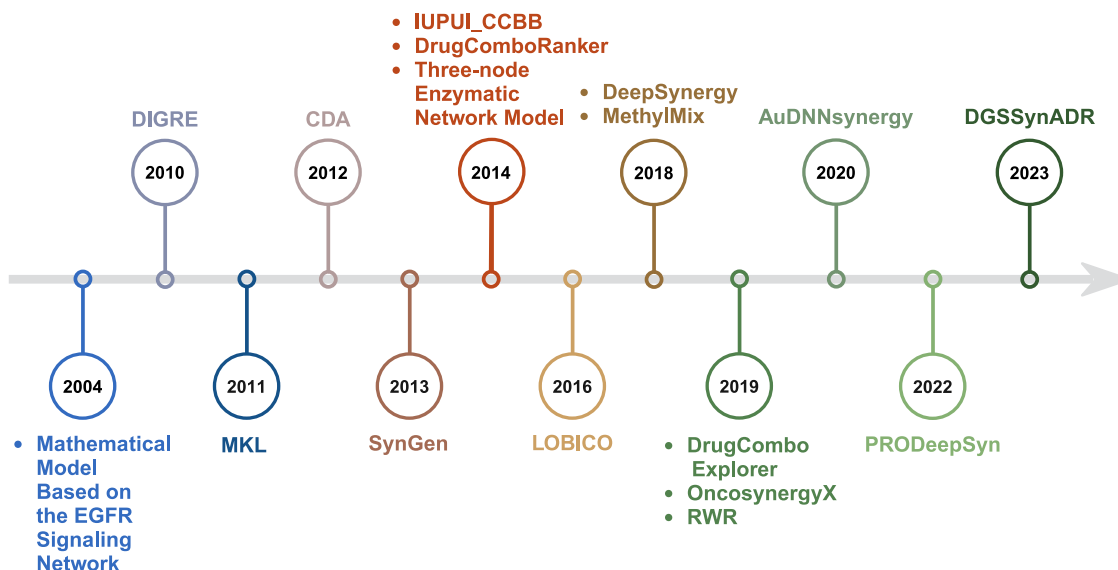


Fig. 1 | Timeline of drug combination prediction models. Chronological development of computational models for drug combination prediction, showing the evolution from traditional machine learning approaches to advanced deep learning

methods over two decades. This figure was created based on the tools provided by Biorender.com (accessed on 7/12/2024).

research. These methods can not only elucidate the biological mechanisms underlying drug synergy but also expedite the discovery and optimization of synergistic combinations. Despite certain advancements, the computational prediction of drug synergistic combinations using multi-omics data still encounters various challenges, including limited mechanistic explanation, unavailable comprehensive expression profiles or drug sensitivity data, and so on⁸. Considering the immense potential and myriad challenges of multi-omics computational methods in drug combination prediction, there is a pressing need for a systematic review and prospective analysis of the current research landscape, key technologies, application cases, and future directions in this field. Nevertheless, there is a paucity of comprehensive review articles summarizing the latest advances and challenges in this domain.

We searched the relevant literature from January 2024 and earlier, identified through PubMed, Web of Science, and Scopus databases using keywords including “synergistic”, “antagonistic”, “drug combinations”, and “multi-omics”. Articles were selected based on their relevance to computational methods and multi-omics data integration. This article will focus on discussing computational models for predicting synergistic drug combinations based on multi-omics data. First, we will summarize the commonality of cooperative antagonistic algorithms and multi-omics integration methods from section “Basic Concepts and Role of AI in Drug Combination Prediction” to “Multi-omics Integration Methods”. Second, we will introduce several representative computational methods that utilize data from various omic dimensions, including genomics, epigenomics, and transcriptomics from section “Single-omics” to section “Integration of Multi-omics Data”, discuss strategies for enhancing the performance of drug synergy prediction models through the integration of multi-omics data in section “Integration of Drug Structure Information or Systems Pharmacology with Multi-omics Data”, and summarize a timeline of the models mentioned in this review (Fig. 1). Finally, we will summarize the primary challenges currently encountered in this field and provide an outlook on potential future research directions in section “Discussion”. Considering the pressing need for precision medicine and personalized therapy, research on drug synergy combination prediction based on multi-omics data holds significant potential for clinical applications and is expected to guide the development of novel synergistic therapies, ultimately benefiting a wider patient population. Key findings and recommendations include: (1) Single omics approaches and integration of multiple omics data types are supposed to be

used according to specific conditions and different diseases; (2) Concerns about the host-gut-intratumoral microbiomics will optimize precision personalized drug combination; (3) Future development should focus on improving model interpretability and clinical validation.

The commonality of cooperative antagonistic algorithms and multi-omics integration methods

Basic concepts and role of AI in drug combination prediction

In combination drug therapy, synergy and antagonism are two fundamental concepts that describe the interactions between drugs. Synergy refers to the phenomenon in which the therapeutic effect of two or more drugs used in combination is greater than the sum of their individual effects when administered separately¹. Conversely, antagonism implies that the combined effect of drugs is less than the sum of their individual therapeutic effects or even lower than the effect of each drug administered independently². Consequently, optimizing multi-drug combinations to maximize synergistic enhancement and minimize antagonistic attenuation is a crucial aspect of combination drug therapy. Artificial intelligence (AI) techniques have introduced novel breakthroughs in the field of drug combination optimization⁶. Cleverly designed AI algorithm frameworks, based on various omics data, can efficiently identify drug combinations with optimal therapeutic effects. Compared to traditional optimization algorithms, these AI-based methods exhibit superior robustness and global optimization capabilities^{6,9}. AI-assisted synergistic and antagonistic drug prediction algorithms have been successfully applied in various fields, including anti-tumor drug screening¹⁰ and antimicrobial drug optimization¹¹, significantly enhancing the efficiency of drug combination optimization. These studies suggest that drug interaction prediction algorithms are poised to become invaluable tools for tackling multi-drug combination challenges.

The commonality of cooperative antagonistic algorithms

Having established the fundamental concepts of drug synergy and antagonism, as well as the transformative role of AI in this field, we now turn our attention to examining the common methodological framework underlying these AI-based prediction algorithms. This framework can be broadly categorized into three key components: data input strategies, feature extraction and selection methods, and validation approaches. Understanding these common elements is crucial as they form the technical

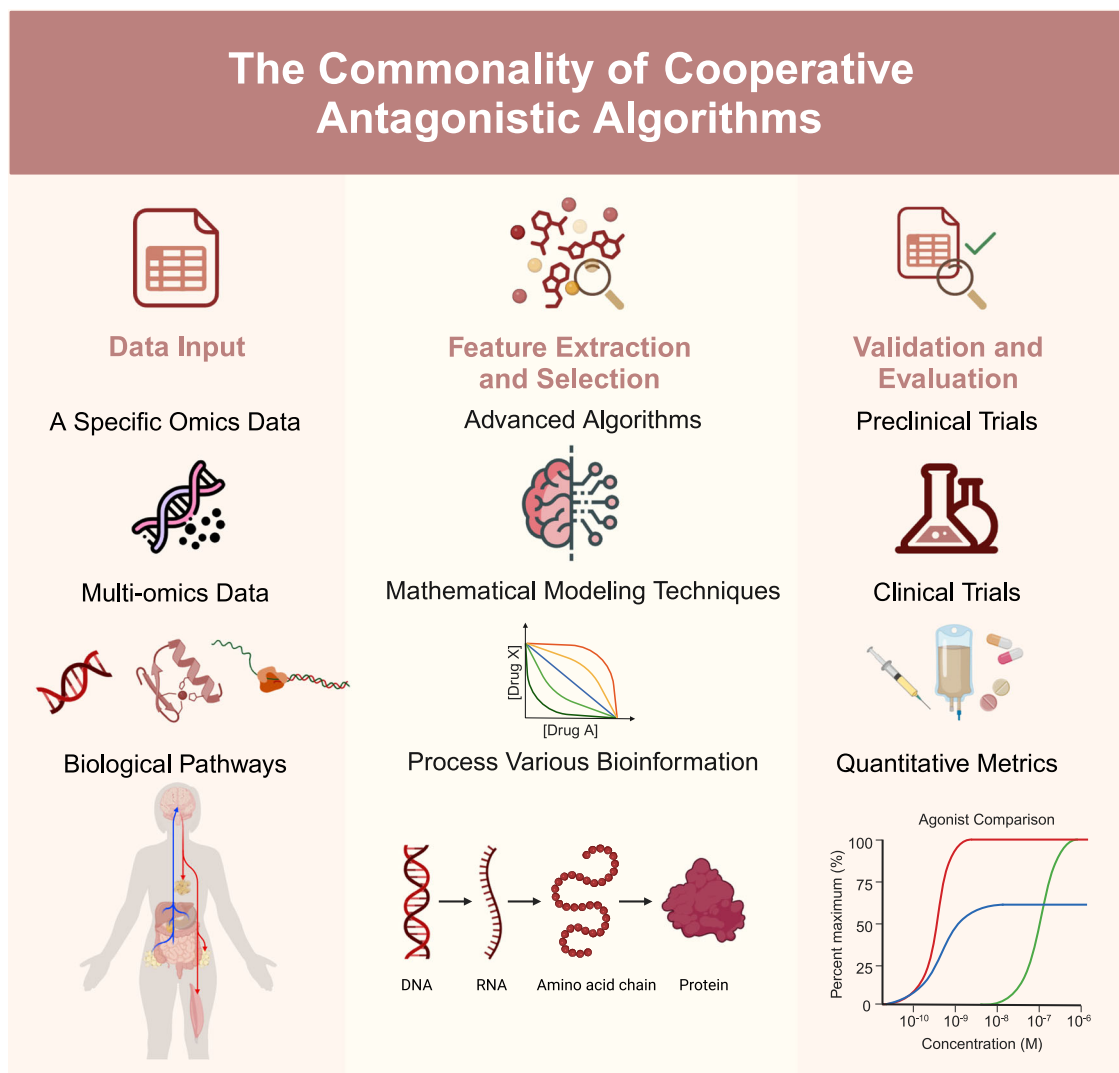


Fig. 2 | Overview of commonly used algorithms for predicting drug synergy and antagonism. The figure delineates the key steps, input data types, and output formats shared by these algorithmic approaches, emphasizing their commonalities in

methodology and application. This figure was created based on the tools provided by Biorender.com (accessed on 4/9/2024).

foundation upon which various AI models are built to predict drug combinations effectively.

Data input. The prediction of drug combinations relies on various types of omics data, each providing unique biological insights (Fig. 2). Genomic data, including gene expression profiles, copy number variations, and mutations, reveal cellular states and potential drug targets. Gene expression data captures the dynamic cellular response to drugs, while copy number variations and mutations help identify genetic alterations that might influence drug sensitivity. These genomic datasets typically require normalization and standardization before use, with expression data often log-transformed and batch effects removed. Proteomic data provides information about protein abundance and post-translational modifications, offering direct insights into drug mechanisms at the protein level. This data usually undergoes intensity normalization and missing value imputation. Pharmacogenomic data links genetic variations to drug responses, requiring careful preprocessing to handle categorical and continuous variables. Additionally, biological pathway information from databases like Kyoto Encyclopedia of Genes and

Genomes¹² helps understand the mechanistic basis of drug interactions^{5,13}, though these networks often need to be converted into appropriate numerical representations for AI models.

These diverse data types can be integrated in three main ways: (1) combining single omics with supplementary multi-omics data, as demonstrated by AuDNNsynergy¹⁴, which primarily uses genomic data while incorporating other omics information; (2) comprehensive multi-omics integration, exemplified by DrugComboExplorer¹⁵, which equally weighs different omics data types to analyze cancer pathways; and (3) network-based integration, where biological pathways and information networks guide the prediction process, as shown in the drug-induced genomic residual effect method, which analyzes transcriptional changes in the context of pathway information.

Feature extraction and selection. Feature extraction and selection are essential preprocessing steps in drug-drug interaction prediction that serve two primary purposes. First, feature extraction transforms raw multi-omics data into meaningful representations that capture the underlying biological patterns. Second, feature selection identifies the most relevant molecular

markers and biological features that contribute to drug responses, thereby reducing data dimensionality and computational complexity. These processes involve converting complex biological information from gene expression, genomics, proteomics, and other omics layers into quantifiable features that can be used by predictive models. Various mathematical and computational approaches, such as Bayesian multi-task multiple kernel learning (MKL) models, facilitate this transformation by extracting key molecular signatures and biological patterns from the multi-dimensional omics data⁵ (Fig. 2). The selected features then serve as inputs for downstream analysis, helping to characterize drug-target interactions and molecular mechanisms underlying drug synergy or antagonism.

Validation and evaluation. Validating predictive models is an indispensable step in the prediction of drug synergy and antagonism. This process typically involves utilizing experimental data from preclinical or clinical studies to validate the efficacy of algorithm-predicted drug combinations. Through this approach, researchers can compare the predicted results with the observed drug responses, verifying the practical applicability and accuracy of the developed algorithms. To evaluate the performance of drug combinations, various quantitative metrics can be employed, including the Bliss Independence (BI) synergy score¹⁶, which is calculated as $S = E_{A+B} - (E_A + E_B)$ where E_{A+B} represents the combined effect of drugs A and B, while E_A and E_B represent their individual effects. A positive S indicates synergy, while a negative S suggests antagonism. The synergy score is a metric specifically designed to quantify the degree to which the effect of two or more drugs is potentiated when administered in combination compared to their individual applications, aiding researchers in understanding the synergistic mechanisms of different drug combinations and their effects in biological systems. Another commonly used metric is the Combination Index (CI)¹⁷:

$CI = (C_{A,x}/IC_{x,A}) + (C_{B,x}/IC_{x,B})$ where $C_{A,x}$ and $C_{B,x}$ are the concentrations of drugs A and B used in combination to achieve $x\%$ effect, and $IC_{x,A}$ and $IC_{x,B}$ are the concentrations required for the same effect when used alone. $CI < 1$ indicates synergy, $CI = 1$ suggests additivity, and $CI > 1$ implies antagonism. These mathematical frameworks provide quantitative assessments of drug interactions, though they may not capture all aspects of complex biological responses. Additional validation measures include comparative analyses with existing drug response data and conducting preclinical studies, such as cell line and animal model experiments, prior to clinical trials. Quantitative metrics can aid researchers in optimizing drug combinations and mitigating uncertainties and risks in clinical trials (Fig. 2). Ultimately, validation through experimental data can not only assess and enhance the accuracy of predictive models but also deepen the understanding of drug action mechanisms, providing a robust scientific foundation for future drug design and therapeutic development.

Multi-omics integration methods

In contemporary pharmaceutical research, the integration of multi-omics data plays a pivotal role. These methods enable the extraction and analysis of data from various biological levels, facilitating more accurate predictions of drug effects and synergies by researchers. These techniques encompass kernel regression, machine learning, graph network approaches, and simulation-based methods, each offering distinct applications and advantages (Fig. 3).

Kernel regression techniques¹³ enable the integration of gene expression and genomic data, enhancing prediction accuracy through the construction of similarity matrices. Machine learning is a computational approach that predicts synergistic effects of drug combinations through multi-omics data, including gene expression data and target information.

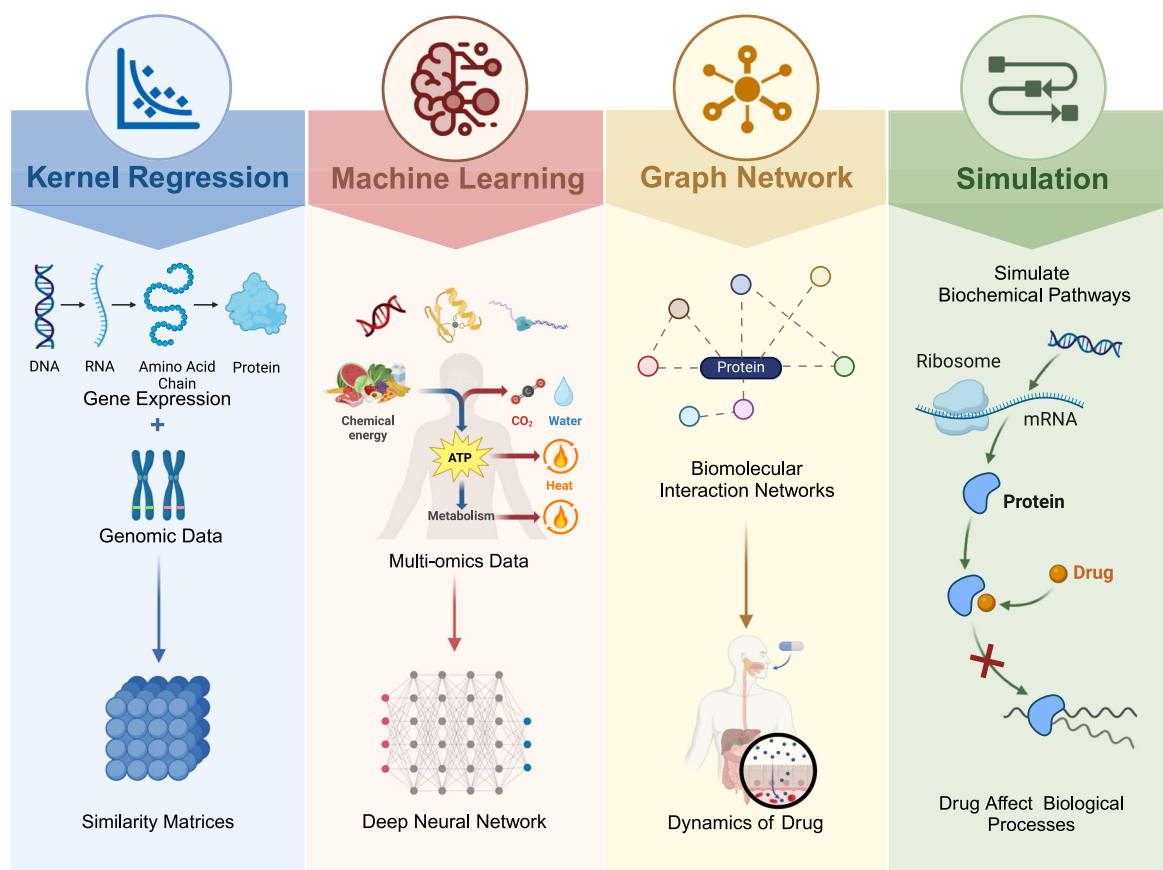


Fig. 3 | Methods of multi-omics data integration. This figure presents various methods for integrating multi-omics data, including kernel regression, machine learning algorithms, graph network approaches, and simulation-based techniques.

Each method offers unique applications and advantages in the context of drug synergy and antagonism prediction. This figure was created based on the tools provided by Biorender.com (accessed on 4/9/2024).

Moreover, graph network approaches harness biomolecular interaction networks, including PPI, providing a sophisticated yet efficient means to investigate the network dynamics of drug effects^{18,19}. Simulation-based methods, including metabolic network models, offer potent tools for elucidating the impact of drugs on specific biological processes by simulating the dynamic alterations in biochemical pathways^{20–22}. These models forecast the potential effects and adverse reactions of drug interventions by meticulously simulating each step in metabolic pathways.

By integrating these diverse techniques, contemporary pharmaceutical research can predict and comprehend drug behavior and efficacy in complex biological systems with greater precision. Employing this multi-dimensional approach, researchers can not only enhance the efficiency and accuracy of drug development but also establish a more robust scientific basis for clinical applications. This strategy of integrating multi-omics data is emerging as a crucial technological pathway for advancing personalized medicine and precision therapy.

To systematically compare and analyze existing drug combination optimization methods, we summarized these approaches across multiple dimensions including implementation characteristics, computational performance, data requirements, and application scope (Tables 1–3). Table 1 provides detailed comparisons of these methods in terms of model type, coverage, consistency, speed, scalability, performance, advantages, and limitations. Table 2 focuses on technical aspects such as implementation challenges, computational costs, data availability issues, validation requirements. Table 3 summarizes these methods from an application perspective, examining input data types, output formats, and applicable disease domains.

These comparative analyses reveal that existing methods primarily rely on multi-omics data, including genomic data, drug chemical features, and protein-protein interaction networks. In terms of output formats, most methods provide quantitative scores or classification predictions for drug combinations. Regarding application scope, while these methods were predominantly developed for cancer treatment, many can be extended to other disease domains with similar data characteristics. Each method exhibits distinct strengths and limitations; selecting an appropriate approach requires comprehensive consideration of specific application scenarios, available data types, and computational resources.

Multi-omics-based predictive modeling of drug synergistic combinations

Single-omics

Genomics-based approaches

DrugComboRanker. DrugComboRanker leverages gene expression profile data and PPI data to predict synergistic drug combinations. DrugComboRanker employs the Bayesian non-negative matrix factorization method to partition the drug functional network into drug communities⁷. The algorithm primarily relies on gene expression profiles from Connectivity Map (CMAP) database²³ (6100 profiles from 4 cancer cell lines treated with 1309 drugs) and drug interaction data from Search Tool for Interactions of Chemicals²⁴ and Biological General Repository for Interaction Datasets²⁵. DrugComboRanker employed drug genomics data ($n \times p$ matrices) and stratified sampling with 60% training and 40% test data. While individual drugs may appear in both sets due to genomics data constraints, all drug combinations in the test set are novel. Drug genomics data is first transformed into low-dimensional representations via Bayesian non-negative matrix factorization, then combined with network features to build drug functional communities. Through these communities, the model reveals potential drug targets and identifies optimal drug combinations by enriching drug targets in complementary modules of the disease signaling network. This approach helps overcome drug resistance issues associated with single drugs. Compared to other methods, DrugComboRanker effectively integrates genomic data of drugs and diseases, enabling a more comprehensive evaluation of how drugs act on dysregulated signaling pathways in diseases. The effectiveness of DrugComboRanker has been validated through case studies of lung adenocarcinoma and breast cancer.

Huang et al. applied DrugComboRanker to evaluate lung adenocarcinoma and endocrine receptor-positive breast cancer, discovering a group of effective drug combinations that ranked at the top of the prediction list⁷. Despite the promising application prospects of DrugComboRanker, it still has some limitations, such as constructing drug networks based solely on cell line data, lacking further preclinical and clinical validation, and not yet integrating certain biomedical knowledge and data, such as drug side effects. In the future, improvements and refinements in these aspects are still needed.

AuDNNsynergy. AuDNNsynergy, a deep learning model, is designed to predict the synergistic effects of drug combinations by integrating multi-omics data. The model utilizes data from The Cancer Genome Atlas (TCGA), which includes gene expression, copy number variations, mutation data, and physicochemical properties of drugs¹⁴. AuDNNsynergy represents a sophisticated approach to drug synergy prediction through its innovative handling of multi-omics data. The model processes three distinct types of molecular data: gene expression profiles as continuous numerical matrices, copy number variations as discrete integer values, and genetic mutations as binary indicators. These heterogeneous data types are first processed through separate autoencoder networks, each optimized for its specific data modality. The gene expression data undergoes normalization and is represented as a real-valued matrix of size $n \times m$, where n represents samples and m represents genes. Copy number variations are encoded as integers indicating deletions or amplifications. Genetic mutation data is transformed into a binary matrix indicating the presence (1) or absence (0) of mutations. The model employs a parallel architecture where each data type is initially processed independently through dedicated autoencoder layers before being integrated at a fusion layer. This architecture enables the model to learn optimal representations from each data type while preserving their unique biological characteristics. The compressed representations are then concatenated and fed into a deep neural network comprising multiple fully connected layers, ultimately producing a continuous synergy score prediction for drug pairs. Notably, the model's training protocol implements a careful data splitting strategy where drug pairs are randomly assigned to training, validation, and test sets while ensuring that cell lines represented in the test set are also present in the training data, enabling the model to learn cell-type specific patterns of drug synergy. The advantage of AuDNNsynergy lies in its comprehensive integration of various data types, enabling more accurate and extensive predictions. By reducing data dimensionality through autoencoders, AuDNNsynergy can effectively handle large datasets and mitigate tissue-specific biases. Moreover, AuDNNsynergy demonstrates superior performance in ranking drug combinations compared to other models, such as DeepSynergy⁶, Random Forests²⁶, and Elastic Nets²⁷. However, AuDNNsynergy also possesses certain limitations. The complex architecture of AuDNNsynergy requires substantial computational resources and storage. The accuracy of the model is largely dependent on the quality and integrity of the input data, which can be challenging if the data is incomplete or biased.

Epigenomics-based approaches

MethylMix. MethylMix is a machine learning model that integrates Deoxyribonucleic Acid (DNA) methylation and gene expression data to predict drug-induced differentially methylated driver genes²⁸. This model has been applied to drug screening for various cancers, including breast and colorectal cancer. MethylMix primarily sources its data from TCGA, automatically downloading and preprocessing both DNA methylation and gene expression datasets across 33 different cancer types in TCGA's database. In epigenomic-based drug synergy prediction, methylation data is represented as beta-value matrices where rows correspond to CpG sites and columns to samples. The model processes these inputs through a three-layer architecture: first, a preprocessing layer handles missing value imputation and batch correction; second, a clustering layer groups correlated CpG probes (correlation threshold 0.7) into functional units; finally, a beta mixture modeling layer identifies differential methylation states. For data

Table 1 | Comparative analysis of drug combination optimization methods: features and performance metrics

Drug-optimizing Combination Methods	Model Type	Coverage	Consistency	Speed	Scalability	Performance	Advantages	Limitations
DrugComboRanker	Genomics-based network computational model	Encompasses 1309 FDA-approved drugs and bioactive compounds; Integrates 6100 gene expression profiles from CMAP database; Includes data from multiple cancer cell lines	Ensures stable community partitioning through Bayesian non-negative matrix factorization; Combines transcriptional response similarity with STITCH database metrics; Maintains model stability using automatic relevance determination priors	Implements efficient multiplicative update algorithms; Employs local majorization-minimization strategy for fast convergence; Enables parallel processing of multiple drug communities	Handles large-scale drug functional networks with 51,164 edges; Supports dynamic expansion of new drug and target data; Allows flexible adjustment of community partition granularity	Validated on lung adenocarcinoma and breast cancer; Successfully identified effective drug combinations that ranked high in predictions; Demonstrated ability to map drug targets on disease signaling networks; Specific quantitative performance metrics were not provided in the excerpts	Effectively integrates genomic data (gene expression profiles) and PPI data; Uses Bayesian non-negative matrix factorization to identify drug communities; Can identify optimal drug combinations by analyzing disease signaling networks; Successfully validated through case studies in lung adenocarcinoma and breast cancer; Helps overcome drug resistance issues associated with single drugs	Relies only on cell line data for drug network construction; Lacks comprehensive preclinical and clinical validation; Does not integrate certain biomedical knowledge like drug side effects; Limited to gene expression and PPI data, missing other potentially valuable omics data types
AuDNNsynergy	Genomics-based deep neural network model	Processes multiple types of Genomic data (gene expression, copy number, genetic mutation)	Employs standardized autoencoder architecture for different data types	Efficient data processing through parallel autoencoder networks	Handle large-scale multi-omics datasets through dimensional reduction	Outperformed DeepSynergy, gradient boosting forests, and elastic nets; Better accuracy in predicting synergistic drug combinations	Integration of multiple types of omics data; Better predictive performance than traditional methods; Cell-line specific predictions	Requires comprehensive multi-omics data; Limited to cancer cell lines with available molecular data; May not generalize well to new drug classes
MethylMix	Genomics-based beta mixture model	Processes data from 33 TCGA cancer types; Handles multiple DNA methylation platforms; Integrates methylation and gene expression data	Implements standardized preprocessing; Applies uniform batch correction; Maintains consistent probe clustering (correlation threshold 0.7)	Enables parallel processing; Automates data downloading and preprocessing	Processes genome-wide methylation data; Handles large-scale multi-omics integration; Accommodates various data formats (arrays and sequencing)	Can identify disease-specific methylation patterns; Successfully identified DNA hypomethylated subtypes in lung squamous cell carcinoma; Effective for cancer subtyping	Automated downloading and preprocessing of TCGA data; Handles missing values and batch effects; Integrates multiple data types	Computationally intensive for large sequencing datasets; Requires matched gene expression data for full functionality; Limited to beta value format data
CDA	Transcriptomics-based computational model	Encompasses 6100 expression profiles from 1309 molecules	Shows reproducible pathway enrichment analysis results across similar drug classes	Performs rapid pattern matching through transcriptional response modules	Handles multiple signaling pathways and drug combinations simultaneously	Successfully identified molecules with similar functions in <i>in silico</i> validation; <i>In vitro</i> validations on non-small cell lung cancer cells and triple-negative breast cancer cells found synergistic effects with certain drug pairs	Targets multiple signaling pathways, potentially increasing efficacy and reducing drug resistance;	May not account for all possible drug interactions or pathways; Requires careful interpretation of results by researchers

Table 1 (continued) | Comparative analysis of drug combination optimization methods: features and performance metrics

Drug-optimizing Combination Methods	Model Type	Coverage	Consistency	Speed	Scalability	Performance	Advantages	Limitations
DIGRE	Transcriptomics-based comparative analysis methodology	Compares four major drug interaction analysis methods; Evaluates synergy/antagonism across different effect levels	Provides similar major conclusions across all methods; Shows reproducible results for synergy detection	Enables rapid visual inspection via curve shift analysis; Allows quick quantitative estimation via combination index; Facilitates efficient comparison through single parameter summary	Handles multiple drug combination ratios; Analyzes large arrays of combinations; Accommodates different effect levels and conditions	Provided consistent major conclusions	Curve shift analysis allows visual data inspection; Isobologram shows optimal concentration ratios; Combination index provides quantitative synergy measures; Universal surface response method summarizes interactions in single parameter	Lacks statistical estimates; Complex interpretation when curves are non-parallel
IUPI_CCBB	Transcriptomics-based computational model	-	-	-	-	Achieved the second-best performance in a challenge-based blind test	Provides a straightforward approach to evaluating the interactions between compounds	Primarily predicts synergistic effects with limited predictive power for antagonistic effects; May require extensive experiments and be costly and time-consuming
SynGen	Transcriptomics-based matrix algebraic model	Screens large-scale drug/gene pairs (>1000 targets); Tests both positive and negative interactions; Applicable across multiple biological systems	Achieves 62–72% synergy detection with only 20% screening; Performance validated across 9 different datasets; Maintains robust prediction accuracy with sparse data (10–20%)	Up to 7-fold faster than traditional screening approaches; Efficient matrix completion algorithm for large-scale predictions; Rapid alternation between sampling and prediction modes	Handles datasets up to several thousand targets; Performance improves with larger target numbers; Computationally efficient for large matrices	Up to 7-fold increase in synergy discovery rate; Can find 75% of synergies by testing only 20% of all possible pairs; Better performance with larger datasets (>1000 targets)	Significantly reduces experimental screening burden; Can incorporate prior biological knowledge; Computationally efficient for large libraries; Applicable to both drug and genetic screens	Performance depends on dataset size; Requires some initial experimental data; Accuracy depends on modularity of interaction network
DeepSynergy	Transcriptomics-based feed-forward neural network	Covers 583 distinct drug combinations; Tests against 39 cancer cell lines from 7 tissue types; Utilizes 38 diverse anticancer drugs (14 experimental, 24 approved)	Achieves 0.73 Pearson correlation between measured and predicted values; Demonstrates 7.2% improvement over second-best method; Shows AUC of 0.90 for novel drug combination classification	Processes large-scale data; Handles high-dimensional input (3984 genomic + 4387 chemical features); Enables efficient exploration of combinatorial space	7.2% improvement over second-best method; Pearson correlation of 0.73 for novel drug combinations; AUC of 0.90 for classification tasks	Works well for predicting novel combinations of known drugs; Only requires basic chemical and genomic information	Poor performance when extrapolating to completely new drugs or cell lines; Limited by size and diversity of training dataset	

Table 1 (continued) | Comparative analysis of drug combination optimization methods: features and performance metrics

Drug-optimizing Combination Methods	Model Type	Coverage	Consistency	Speed	Scalability	Performance	Advantages	Limitations
PRODeepSyn	Protein omics-based deep learning model	Covers 583 drug combinations across 39 cancer cell lines from 7 tissues; PPI networks, gene expression, and mutation data	Achieves highest Pearson correlation coefficient of 0.75; Shows lowest root mean square error of 15.08; Demonstrates consistency with previous studies in case analysis	Employs efficient dimensionality reduction; Uses batch normalization to accelerate training	Handles large-scale datasets; Processes complex network data	RMSE: 15.08 (lowest); Pearson correlation coefficient: 0.75 (highest); AUC-Receiver Operating Characteristic (ROC): 0.90; Cohen's Kappa: 0.53	Integrates PPI network information for cell line feature construction; Capable of handling complex nonlinear biological processes	Validated only on in vitro data; Requires complete omics data and PPI network data
DGSSynADR	Protein omics-based deep learning model	Encompasses 52 drugs; Processes 80,877 drug combinations; Analyzes 60 cancer cell lines	Integrates 6 synergy metrics; Benchmarks against 7 competing methods	Reduces computation complexity via an graph convolutional network; Enables end-to-end training efficiency	Constructs heterogeneous graph networks; Handles large-scale datasets	Outperformed seven competing methods	Reduced graph computation complexity; Improved feature representation through global weighted aggregation; Effective dimension transformation	Limited to 52 drugs due to data intersection constraints; Requires multiple types of biological data availability
LOBICO	Multi-omics-based logistic regression model	Maps complex biological interactions	Maps complex biological interactions	Solves complex computations	Handles large datasets effectively	Improved accuracy in predicting drug sensitivity in cancer cell lines	Retains continuous output information, allowing for robust and precise logic modeling; Adjustable parameters enable customization for desired sensitivity and specificity levels, making it suitable for clinical applications	Tissue specificity among different cancer types and model complexity are recognized challenges
MKL	Multi-omics-based multiple kernel discriminative model	Spans linear, nonlinear and data-dependent kernel combinations, accommodating diverse similarity measures and multiple data sources	Demonstrates theoretical convergence properties for various combination methods, particularly for convex optimization approaches	Fixed rules and heuristic methods are fast, while optimization-based approaches require more computational resources	Shows limitations with large-scale problems, especially for methods requiring complex optimization procedures	Generally performs better than single kernel methods; No large accuracy improvements over best single kernel; More robust than selecting single kernel	Combines multiple data sources/representations; Automatic kernel selection/combinatorial; More flexible than single kernel approaches; Capture different notions of similarity	Higher computational complexity than single kernel; May not provide significant accuracy gains; Training time can be substantial; Model complexity increases with number of kernels
DrugComboExplorer	Multi-omics-based network-based computational approach	Integrates drugs/compounds pharmacogenomics data; Processes multiple omics data types	Validates predictions through cell viability assays; higher prediction accuracy compared to existing approaches; Shows reliable prioritization of synergistic combinations	Employs efficient algorithms; Enables large-scale drug combination screening; Provides rapid network-based efficacy assessment	Handles vast drug combinatory space; Processes multi-dimensional omics datasets; Accommodates growing pharmacogenomics databases	Evaluated on lung adenocarcinoma and breast cancer; Successfully identified effective drug combinations	Systematic approach for large-scale drug combination screening; Integrates multiple data types; Focuses on FDA-approved drugs and bioactive compounds; Provides mechanistic insights	Requires extensive genomic profiling data; Limited to drugs with available genomic profiles

Table 1 (continued) | Comparative analysis of drug combination optimization methods: features and performance metrics

Drug-optimizing Combination Methods	Model Type	Coverage	Consistency	Speed	Scalability	Performance	Advantages	Limitations
OncosymeryX	Multi-omics-based XGBoost machine learning model	Processes large-scale multi-omics data including somatic mutations, copy-number alterations, DNA methylation, gene expression profiles and compound information	Achieves consistent performance with mean primary metric of 0.39 across ten repetitions of tenfold cross-validation	Utilizes optimized XGBoost parameters for efficient training and prediction	Handles features in full feature set or features in abridged set across multiple cell lines and drugs	Primary metric = 0.36; Tie-breaker metric = 0.37; Mean primary metric of 0.39 with 10-fold cross-validation; Ranked in top 15 out of 76 submissions	Uses comprehensive biological feature set; Integrates multiple data types; Reduces overfitting through biologically informed feature curation	Relies heavily on gene expression and monotherapy data; Performance drops when excluding gene expression/monotherapy data; IC50 as sensitivity measure may not be optimal
RWR	Integrated heterogeneous signaling pathway model	Explores multiple interaction networks simultaneously including protein-protein, co-expression, and phenotype networks	Provides stable results through iteration until convergence	Converges efficiently using restart probability of 0.7	Handles multiplex networks with multiple layers and heterogeneous networks with different node types	Outperformed previous RWR approaches in disease gene prediction; Validated through leave-one-out cross-validation	Integrate multiple types of biological networks; Leverages complementary interaction data; Maintains individual network features and topologies	Requires restart probability parameter tuning; Depends on quality and completeness of input networks
Mathematical Model Based on the EGFR Signaling Network	Integrating signaling network deterministic kinetic model	Captures receptor activation, dimerization, phosphorylation and signal propagation	Based on established EGFR network topology; Uses empirically-derived rate constants; Incorporates mechanistic understanding of receptor regulation	Non-stiff system of ordinary differential equations solvable by standard numerical methods; Fourth-order Runge-Kutta algorithm sufficient for integration; Computationally tractable for parameter exploration	Framework extendable to additional pathway components; Can incorporate different types of inhibitors; Allows analysis of multiple drug combinations	Successfully simulates EGFR network dynamics; Demonstrates enhanced signal attenuation with multiple inhibition points; Shows synergistic effects of targeting serial processes	Captures complex network interactions; Predicts combination therapy outcomes; Non-stiff system, computationally tractable	Simplified network topology; Limited feedback mechanisms; Does not include spatial aspects; No receptor trafficking dynamics
Three-node Enzymatic Network Model	Mathematical model of three interacting enzymes using ordinary differential equations Based on Michaelis-Menten kinetics	Models all possible 3-node enzymatic network topologies; Studies two-drug combinations through target interactions Examines synergistic and antagonistic drug effects	Results largely independent of parameter variations; Most combinations showed consistent behavior across parameter sets	Solved nonlinear equations; Processed a large number of drug combinations; Each network tested with 100,000 parameter sets	Limited to three-node networks; Successfully calculated over 100 cases for each studied combination; Can be extended as simplified versions of more complex networks	Successfully identified consistent synergistic and antagonistic motifs; Results largely independent of parameter variations; Solved over 33,798 drug combinations	Simple but informative representation of drug target interactions; Parameter-independent predictions; Comprehensive catalog of drug interaction patterns	Only considers three-node networks; Limited to enzymatic interactions; Excludes pharmacokinetic drug interactions; Requires stable steady states

Table 2 | Implementation and technical challenges in drug combination optimization methods

Drug-optimizing Combination Methods	Implementation Challenges	Computational Costs	Data Availability Issues	Validation Requirements
DrugComboRanker	Relies only on cell line data for drug network construction Lacks comprehensive preclinical and clinical validation; Does not integrate certain biomedical knowledge like drug side effects; Limited to gene expression and PPI data, missing other potentially valuable omics data types	Processes large datasets (6100 profiles from 4 cancer cell lines with 1309 drugs); Requires matrix factorization computations; Needs to analyze complex biological networks; Must handle multiple databases simultaneously	Relies on access to CMAP database; Requires drug interaction data from multiple sources (STITCH, BioGRID); Needs updated PPI network information; Limited by available cell line data	Needs experimental validation of predicted drug combinations; Requires clinical validation beyond cell line studies; Should validate predictions across different disease contexts; Needs comparison with other prediction methods
AuDNNsynergy	Need for high quality multi-omics data; Complex data preprocessing requirements; Model training requires significant computational resources	Requires substantial computational resources for training due to: Processing of high-dimensional multi-omics data; Deep neural network architecture; Multiple cell line-specific models	Requires comprehensive multi-omics profiles for cell lines; Data quality and consistency across different omics; platforms Need for standardized drug combination screening data	Experimental validation of predicted synergistic combinations; Cross-validation across different cell lines; Independent test sets for model evaluation; Need for replication in clinical settings or animal models
MethyMix	Handling missing values; Batch effect correction; Computational demands for large datasets	Higher computational demands for sequencing data; Parallel processing support to improve performance; Mixture modeling is computationally demanding	Requires complete datasets from both normal and cancer samples; Needs matched gene expression data; Dependent on TCGA data availability	Statistical validation through consensus clustering; Biological validation through pathway analysis; Clinical validation of identified subtypes
CDA	Extracting biologically meaningful information from genome-wide expression analysis; Standardizing bioinformatics approaches for signature extraction; Integrating data from multiple public databases for a unified analysis	May have significant computational requirements	Reliance on public databases for gene expression data and pathway information	In vitro validations are necessary
DIGRE	Requires specialized software/programming; Complex mathematical modeling; Need for careful experimental design	Appears moderate given mathematical complexity	Requires extensive experimental data points; Multiple drug combinations and concentrations needed	Statistical validation needed for some methods; Reproducibility testing; Comparison across multiple methods recommended
IUPUI_CCBB	Without considering other factors such as drug metabolism, transport, and target binding may impact predictive accuracy	May have moderate computational requirements.	Requires gene expression profile data following treatment with highly toxic compounds, which may be limited or difficult to obtain	Requires rigorous experimental validation and clinical trials
SynGen	Uses nuclear norm constraints; Alternates between propensity-based sampling and prediction-driven screening; Can handle datasets up to several thousand targets	Not specifically stated	Not specifically stated	Not specifically stated
DeepSynergy	Requires double representation of drug pairs to ensure order-invariant predictions; Needs careful normalization strategy for heterogeneous input data	Not specifically stated	Available at https://www.bioinf.jku.at/software/DeepSynergy/	Nested cross-validation; Separate validation for known vs novel drug combinations
PRODeepSyn	Integration of heterogeneous data sources; Processing of high-dimensional sparse omics data	Not specifically stated	O'Neil dataset	Experimental validation of predictions needed; Cross-validation required for assessing prediction ability on novel drug combinations

Table 2 (continued) | Implementation and technical challenges in drug combination optimization methods

Drug-optimizing Methods	Implementation Challenges	Computational Costs	Data Availability Issues	Validation Requirements
DGSSynADR	Integration of heterogeneous biological data; Construction of complex graph networks; End-to-end training requirements	Not specifically stated	Available at https://github.com/DHUBlab/DGSSynADR	Validation through previous findings in case studies; Ablation studies for component validation
LOBICO	Optimizes the algorithm for better interpretability and practicality	May require moderate computational requirements	Requires comprehensive genomic and transcriptomic data	May need rigorous experimental validation and clinical trials
MKL	Choosing appropriate kernel combination method; Optimization of kernel weights; Computational scalability; Parameter selection	Higher than single kernel methods	Requires multiple feature representations; Training data with labels needed; Validation data for parameter tuning	Cross-validation for parameter selection; Performance validation on test sets; Kernel weight validation; Model selection criteria
DrugComboExplorer	Complex network construction and analysis; Integration of heterogeneous data types; Parameter optimization for matrix factorization	Not specifically stated	Program available upon request; Requires access to CMAP and protein interaction databases	Experimental validation of predicted drug combinations; Clinical validation for practical application
OncosynergyX	Large feature set requires significant processing	Not specifically stated	Not specifically stated	Experimental validation of predicted synergistic combinations; Clinical trials for promising drug pairs
RWR	Handling large-scale matrices; Memory requirements	Iterative matrix computations; Scale depends on network size	Available at https://github.com/alberto-valdeolivas/RWR-MH	Leave-one-out cross-validation; Known disease-gene associations as ground truth
Mathematical Model Based on the EGFR Signaling Network	Parameter estimation accuracy; Initial condition sensitivity; Network topology completeness	Moderate (solvable via 4th order Runge-Kutta method); Non-stiff system requires standard numerical methods	Rate constants from literature; Initial conditions based on cell culture studies	Experimental validation of predicted combination effects; Verification of synergy predictions; Clinical correlation
Three-node Enzymatic Network Model	Solving nonlinear equations for steady states; Ensuring system stability; Computing dose-response relationships within physiological ranges	Required 100,000 parameter sets per network; Multiple numerical solutions needed per parameter set	Model equations and parameters provided in paper; No external data dependencies	Experimental validation needed; Translation to larger networks unclear; Biological relevance of simplified networks

Table 3 | Data characteristics and disease applications of drug combination methods

Drug-optimizing Combination Methods	Input Data	Output Data	Diseases (Initially and Primary Focus)
DrugComboRanker	Drug genomic profiles from CMAP (6100 gene expression profiles); Patient genomic profiles; Protein interactome data; Drug similarity data from STITCH database	Prioritized drug combinations; Drug target mapping on disease signaling networks; Drug functional network communities	Cancers; Other diseases with similar input data characteristics
AuDNNsynergy	Multi-omics data from cancer cell lines: Gene expression; Copy number variation; Genetic mutations	Binary classification of drug pair synergy (synergistic vs non-synergistic)	Cancers; Other diseases with similar input data characteristics
MethylMix	DNA methylation data; Matched gene expression data	Differential Methylation Values; Transcriptionally predictive methylation states; Identification of hyper/hypomethylated genes; DNA methylation subtypes through consensus clustering	Lung squamous cell carcinoma; Head and neck squamous cell carcinoma; Ovarian cancer; Other diseases with similar input data characteristics
CDA	Gene expression profiles from cell lines	Lists of single drugs and combinatorial drugs showing similar expression patterns; Predictions of drug combinations that target multiple signaling pathways	Non-small cell lung cancer; Triple-negative breast cancer; Other diseases with similar input data characteristics
DIGRE	Drug concentration-effect data from cell culture experiments; Drug combination ratios	Quantitative measures of drug synergy/antagonism; Interaction parameters; Statistical estimates of synergy vs additivity	Ileocecal adenocarcinoma; Other diseases with similar input data characteristics
IUPUI_CCBB	Gene expression profiles	Predictions of synergistic or antagonistic effects between drug combinations	Various diseases, particularly in cancer treatment; Other diseases with similar input data characteristics
SynGen	Small set of experimentally determined target pairs; Pre-existing similarity data (gene ontology, protein-protein interactions); Drug-drug or gene-gene interaction scores	Predictions of synergistic target pairs; Interaction scores for untested pairs	Glioblastoma; Lung cancer; Colon cancer; Other diseases with similar input data characteristics
PRODeepSyn	Molecular fingerprints and descriptors; Gene expression and mutation data; Protein-protein interaction network	Synergy scores for drug combinations	Cancers; Other diseases with similar input data characteristics
DGSSynADR	Drug chemical structures and features; Drug-target interactions; Protein-protein interactions; Cell line genomic data; Multi-omics data	Synergy scores for drug combinations in different cancer cell lines	Cancers; Other diseases with similar input data characteristics
DeepSynergy	Chemical descriptors of drug pairs; Gene expression profiles of cancer cell lines	Predicted drug synergy scores based on Loewe Additivity	Cancers; Other diseases with similar input data characteristics
LOBICO	Multiple molecular data types, including gene mutations, copy number variations, DNA methylation, and gene expression data	Predicted continuous drug response	Cancers; Other diseases with similar input data characteristics
MKL	Multiple feature representations/sources; Different similarity measures/kernel functions; Training data with class labels	Combined kernel function; Classification/prediction results; Kernel weights	Diseases with similar input data characteristics
DrugComboExplorer	Drug genomic profiles from CMAP database; Disease genomic profiles; Protein-protein interaction data; Known drug-target interactions	Ranked list of synergistic drug combinations; Mapped drug targets on disease signaling networks; Mechanisms of action for drug combinations	Lung adenocarcinoma; Endocrine receptor positive breast cancer; Other diseases with similar input data characteristics
OncosynergyX	Drug features (targets, chemical/biological properties); Genomics; Epigenomics; Transcriptomics	Predicted drug synergy scores for drug combinations	Cancers; Other diseases with similar input data characteristics
RWR	Multiplex biological networks (protein-protein interactions, pathways, co-expression); Disease-disease similarity networks; Known gene-disease associations	Ranked list of genes based on proximity to seed nodes; Proximity scores between genes and diseases	Wiedemann-Rautenstrauch syndrome; SHORT syndrome; Other genetic diseases with known associated genes; Other diseases with similar input data characteristics

Table 3 (continued) | Data characteristics and disease applications of drug combination methods

Drug-optimizing Combination Methods	Input Data	Output Data	Diseases (Initially and Primary Focus)
Mathematical Model Based on the EGFR Signaling Network	Initial concentrations of proteins and molecules; 50 reaction rate constants; Inhibition parameters	Time-evolution of 23 protein/molecule concentrations; Signal attenuation profiles under different inhibition scenarios	Cancers, specifically targeting EGFR-dependent tumors; Other diseases with similar input data characteristics
Three-node Enzymatic Network Model	Network topology (activation/inhibition relationships between nodes); Kinetic parameters (KM: 0.001–10 μ M, kcat: 0.1–10 s^{-1}); Drug inhibition parameters; Fixed concentrations	Steady-state concentrations; Drug combination effects (synergistic/antagonistic/additive); Combination Index (CI) values; Dose-response relationships (isobolograms)	Cancer (particularly MAPK and PI3K/Akt/mTOR pathway targeting); Diseases treated with enzyme inhibitors; Other diseases with similar input data characteristics

splitting, a stratified k-fold cross-validation approach is employed where methylation profiles are partitioned to ensure both training and test sets contain representative samples from each methylation state. This method accounts for the continuous nature of methylation data while maintaining biological relevance in the prediction task. The model outputs differential methylation values that quantify the degree of abnormal methylation, which can then be used to predict drug synergy scores through correlation with treatment response data. This approach enables the prediction of drug efficacy and mechanisms of action. The advantages of such a model include its intuitiveness, ease of understanding, and high interpretability. Furthermore, the main output is “Differential Methylation values” that can be used for cancer subtyping. However, the limitations of these models include the lack of consideration for the interactive regulation between methylation and other epigenetic modifications, as well as the susceptibility of differential analysis methods to confounding factors such as batch effects.

Transcriptomics-based approaches

Combinatorial Drug Assembler (CDA). CDA is an innovative technique that utilizes genomics and bioinformatics approaches to discover drug combinations targeting multiple signaling pathways²⁹. CDA sources its data primarily from CMAP²³, containing 6100 expression profiles representing 1309 molecules tested on five different human cancer cell lines, along with pathway gene set data from the Pathway Interaction Database (<http://pid.nci.nih.gov/>) comprising 166 pathways and 2297 genes. In the CDA model, transcriptomic inputs are processed through a two-stage architecture: first, gene set enrichment analysis identifies relevant signaling pathways, then pattern-matching algorithms assess expression similarities between drug treatments. The integration of multiple drugs’ transcriptional responses can follow either a concatenation approach, where expression profiles are combined before analysis, or a pathway-specific fusion strategy, where drug effects are evaluated independently within each signaling module before integration. The CDA evaluation demonstrates the highest predictive power when at least one drug in test pairs appears in the training set, allowing the model to leverage known transcriptional responses. However, the model can also make predictions for novel drug pairs by comparing their pathway-level effects to known patterns. This balance between leveraging existing knowledge and generalizing to new combinations is particularly important when working with transcriptomic data, as expression patterns can capture both direct drug effects and downstream pathway perturbations. One significant advantage of CDA lies in its ability to enhance efficacy and reduce toxicity by identifying synergistic drug combinations that allow for the use of lower dosages. Moreover, by simultaneously targeting multiple pathways, CDA can effectively delay or prevent the emergence of drug resistance, a common issue with single-pathway-targeted therapies³⁰. Furthermore, CDA employs high-throughput and data integration approaches, enabling the rapid and efficient identification of potential drug combinations¹³. In practical applications, CDA has demonstrated immense potential. It can be applied in the field of personalized medicine to identify the most effective drug combinations for individuals based on their specific gene expression profiles³⁰. CDA has demonstrated positive results in the treatment of various

cancer types, such as non-small cell lung cancer and triple-negative breast cancer, offering patients novel treatment options^{13,30}. Additionally, CDA has been utilized to identify new estrogen antagonists and discover novel applications for existing drugs, greatly expanding the scope of drug discovery³⁰. Although CDA is a highly promising technology, it also has some limitations. The success of CDA largely depends on the accuracy of gene expression data and the ability to interpret complex biological data, posing significant challenges for researchers. Simultaneously, the effectiveness of CDA systems is also limited by the availability and quality of existing molecular and pharmacological data. Furthermore, the predictions made by CDA are constrained by the scientific understanding of disease mechanisms and drug interactions.

Drug-Induced Genomic Residual Effect (DIGRE). DIGRE is a computational approach employed to forecast whether drug combinations display synergistic or antagonistic effects^{13,31}. DIGRE utilizes gene expression data obtained from treating human B cells with 14 individual compounds at various time points and concentrations, combined with dose–response curves and baseline genetic profiles³¹. The mathematical models for analyzing drug–drug interactions require precise quantitative inputs. In this study, experimental data inputs include drug concentrations as real values and drug effects measured as percentage values. For methods like combination index analysis, the data inputs were fitted to the Hill equation without logarithmic transformation using nonlinear regression. The outputs vary by method—*isobologram* provides graphical plots indicating synergy/antagonism regions, while the combination index generates numerical scores quantifying interaction extent. Model validation typically involves analyzing multiple drug concentration ratios under different conditions to ensure robust conclusions about drug interaction patterns. DIGRE can yield statistically significant predictions of synergistic or antagonistic drug effects, signifying the viability of employing computational approaches for prediction. In practical applications, DIGRE can complement expensive high-throughput combinatorial screening experiments by prioritizing combinations for experimental validation¹³. Nevertheless, the current predictive accuracy and robustness of the DIGRE approach necessitate further enhancement. Drug action mechanisms are intricate and multifaceted, and solely considering transcriptomic alterations may not comprehensively reflect the effects of drugs. Furthermore, data variations across diverse cell lines and experimental conditions may also influence the reliability of the prediction results. To surmount these limitations, future studies should integrate multi-omics data to construct more comprehensive drug action models. Concurrently, augmenting the diversity and scale of training datasets is essential to enhance the generalizability of the models. Moreover, the drug combinations predicted by DIGRE still necessitate rigorous experimental validation and clinical trials to ensure their safety and efficacy. While computational approaches can offer valuable predictions and guidance, they cannot completely replace conventional drug development processes. How to fully leverage the advantages of computational approaches such as DIGRE in clinical translation while closely integrating them with experimental research warrants further exploration.

IUPUI_CCBB. The IUPUI_CCBB method utilized gene expression data as its only data source, and computed a Pearson correlation between gene expression profiles of compound pairs using genes that were differentially expressed in at least one compound treatment¹³. The IUPUI_CCBB method is predicated on the hypothesis that the activity of compounds can be estimated through their impact on significantly differentially expressed genes following treatment with highly toxic compounds¹³. IUPUI_CCBB's method utilized gene expression profiles as primary input data, specifically analyzing the transcriptomic response patterns from cells treated with individual compounds at multiple time points and concentrations. The method ranked compound pairs based on their likelihood of synergistic interaction by examining the statistical significance of differentially expressed genes in response to single compound treatments. The approach focused on identifying a core set of genes that showed significant differential expression in at least one compound treatment condition, then calculated interaction scores. This targeted analysis of transcriptional responses helped predict whether compound pairs would exhibit synergistic or antagonistic effects when combined. A key advantage of the IUPUI_CCBB method resides in its emphasis on the capability to predict synergistic effects of compounds. By analyzing the expression alterations of crucial genes elicited by highly toxic compounds, the method strives to capture information pertaining to the mechanisms of synergistic effects. Moreover, by juxtaposing the effects of two compounds on these crucial genes, the IUPUI_CCBB method offers a straightforward approach to evaluating the interactions between compounds¹³. In practical applications, the IUPUI_CCBB method attained the second-best performance in a challenge-based blind test, demonstrating its robust capability in predicting the synergistic effects of compounds³². This stringent performance evaluation approach affords an unbiased measure of the method's predictive power and enables comparisons with other methods. The IUPUI_CCBB method predominantly focuses on predicting synergistic effects of compounds, whereas its predictive power for antagonistic effects is restricted. This may constrain the method's application in identifying antagonistic drug combinations that ought to be avoided. The IUPUI_CCBB method is reliant on gene expression profile data following treatment with highly toxic compounds³³. Obtaining such data may necessitate conducting extensive experiments, which can be costly and time-consuming. This may restrict the method's application in large-scale drug combination screening. Furthermore, the method predominantly relies on gene expression alterations to predict drug interactions without considering other factors that may influence drug interactions, such as drug metabolism, transport, and target binding. This may impact the predictive accuracy of the method. Moreover, the performance evaluation of this method is predominantly based on a single challenge dataset^{13,34,35}. Although this affords an unbiased assessment, it may not fully reflect the method's performance on other datasets or in real-world applications. Validation of more diverse types of datasets is imperative.

SynGen. Analyzing drug–drug interaction data from cancer cell lines and gene–gene interaction data from yeast screens, SynGen implements a novel framework that combines matrix algebra with experimental validation to identify functional target pairs³⁶. The SynGen method postulates that compound combinations with the capacity to synergistically disrupt these master regulator (MR) activities may demonstrate synergistic effects³⁶. The SynGen methodology employs a sophisticated approach to predict drug synergies by analyzing MR activities through multi-dimensional data integration. The algorithm processes gene expression profiles following drug treatment into numerical vectors representing MR activity patterns. These patterns are transformed into real-valued matrices that capture the regulatory relationships between compounds and MRs. The method then applies an innovative matrix completion technique combining three key components: experimental data constraints, modularity requirements, and target similarity information. This integrated approach enables efficient identification of complementary drug pairs that could synergistically modulate MR activities, significantly outperforming traditional screening

methods in both accuracy and efficiency. The strength of the SynGen method resides in its provision of a mechanism-based approach to forecast compound synergy. By concentrating on MR activity patterns, the method endeavors to capture the pivotal regulatory factors governing cellular phenotypes. Furthermore, by discerning compound pairs that can maximally perturb or enhance these MRs, SynGen presents a straightforward approach to evaluate compound interactions³³. The SynGen method epitomizes an inventive mechanism-based approach for forecasting compound synergy and affords a basis for future enhancements and extensions. In pragmatic applications, researchers utilize predefined experimental datasets to corroborate SynGen's predictive performance, and the findings suggest that the method manifests high sensitivity in forecasting synergistic effects³². Nevertheless, the SynGen method performs inadequately in forecasting antagonistic effects, which may constrain its efficacy in specific applications³². Moreover, the method is contingent upon precisely deducing MR activity patterns, which can be demanding in certain instances¹³.

DeepSynergy. DeepSynergy is a deep learning-based computational method that leverages gene expression profiles and drug property data to predict the synergistic effects of drug combinations⁶. Building upon more than 20,000 anti-cancer drug synergy measurements from a high-throughput screening study, DeepSynergy processes data from 38 drugs tested against 39 cancer cell lines⁶. DeepSynergy employs a structured multi-modal input architecture where drug features are represented as high-dimensional vectors combining binary toxicophore features, chemical footprints, and physicochemical properties. Cell line features are encoded as real-valued gene expression vectors. The model uses feed-forward layers with a conic architecture, where input vectors are concatenated and processed through 2–3 hidden layers with decreasing neuron counts. Regarding data splitting, DeepSynergy performs best when predicting novel combinations of known drugs/cell lines, while extrapolation to completely novel drugs shows limited performance due to dataset constraints. The strength of DeepSynergy resides in its robust feature learning and abstraction capabilities, which enable the model to capture the intricate nonlinear relationships between gene expression and drug properties, consequently enhancing prediction performance. DeepSynergy has exhibited promising potential for application in predicting synergistic combinations of anticancer drugs across various studies. A study underscores the potential of DeepSynergy as a tool for precision medicine in cancer, providing novel insights for developing personalized drug combination therapy strategies⁶. However, DeepSynergy also has several limitations. First, the model's interpretability is limited, rendering it challenging to provide insights into the biological mechanisms underlying drug synergy. Second, the model's performance is dependent on a substantial amount of high-quality training data, which may present acquisition challenges. Furthermore, the model's generalizability may vary across different cancer types and cell lines, necessitating further validation.

Protein omics-based approaches

PRODeepSyn. PRODeepSyn is an innovative deep-learning approach that effectively predicts synergistic effects of drug combinations by integrating protein expression data and chemical structure information³⁷. PRODeepSyn leverages three main data sources: the O'Neil drug synergy dataset for drug combinations and their effects, multi-omics data (gene expression and mutation data) for cell lines, and protein–protein interaction networks from the STRING database³⁷. PRODeepSyn implements a sophisticated multi-omics integration approach using PPI networks containing interactions between proteins. The model architecture employs two-layer graph convolutional networks for network processing, followed by feed-forward neural networks with batch normalization for prediction. Data splitting ensures drug combinations do not overlap between folds, enabling evaluation of the model's ability to predict novel drug combinations through cross-validation. This model showcases significant strengths, notably its innovative integration of complex biological data to enhance prediction

accuracy. However, it also faces challenges, such as the complexity and computational demands needed to handle and process large multi-omic datasets and intricate network structures.

DGSSynADR. DGSSynADR is a computational method that integrates deep learning and heterogeneous network analysis to predict the synergistic combinatorial effects of anticancer drugs³⁸. Drawing from DrugComb³⁹, Chemical Database, and Enzymes of the BioMolecular General Repository Laboratory⁴⁰, HuRI⁴¹, and Cancer Cell Line Encyclopedia⁴², DGSSynADR leverages comprehensive drug interaction data and multi-omics information to power its predictions³⁸. The DGSSynADR model integrates multiple data modalities through a heterogeneous graph structure. The input features include binary drug fingerprints, drug-target interactions, protein-protein interactions, and gene expression profiles. These data are first processed through a message passing neural network that maps different types of information into low-dimensional embedding vectors. The model architecture employs a graph convolutional network for feature extraction, followed by a bilinear predictor that fuses drug pair and cell line features. The output consists of six different synergy scores. DGSSynADR offers several advantages, including the integration of multi-omics data and heterogeneous networks, the utilization of deep learning methods for automatic key feature extraction, and the introduction of attention mechanisms and low-rank representations to enhance generalization ability and robustness¹⁹. Across multiple cancer types and datasets, DGSSynADR has exhibited superior predictive performance compared to traditional machine learning methods. Nevertheless, DGSSynADR also encounters challenges in practical applications, including data quality, computational complexity, and clinical validation. Despite its limitations, DGSSynADR demonstrates extensive application prospects in anticancer drug combination prediction and precision medicine. Through predictive analysis, DGSSynADR can facilitate the identification of novel drug combinations with a high likelihood of producing synergistic effects, thus guiding experimental validation and clinical trials. By integrating patients' molecular characteristics, DGSSynADR can predict the optimal personalized drug combination regimens, thereby promoting the development of precision medicine. In the context of drug action mechanism research and clinical decision support, DGSSynADR is also anticipated to play a pivotal role. However, the DGSSynADR model is not without limitations. Firstly, its performance is contingent upon the quality and comprehensiveness of the available biological data, which may not always be exhaustive or representative of all patient variability. Secondly, while the model excels in predicting synergistic effects, it may require further refinement and validation with larger and more diverse clinical datasets to ensure its generalizability and accuracy across different populations and cancer types. Despite these considerations, DGSSynADR offers a promising approach to enhance our understanding of drug interactions and to advance the field of personalized oncology. In the future, through multidisciplinary collaboration, ongoing method refinements, data scale expansion, and enhanced clinical translation, DGSSynADR is poised to deliver personalized treatment options to a broader spectrum of cancer patients.

Integration of multi-omics data

Logic Optimization for Binary Input to Continuous Output (LOBICO).

The LOBICO algorithm is an innovative approach that integrates multiple molecular data types, including gene mutations, copy number variations, DNA methylation, and gene expression, to create a logistic regression model⁴³. LOBICO was applied to a panel of 714 cancer cell lines, using binary mutation status of 60 features (54 cancer genes plus 6 gene fusions) as input to predict continuous drug response across 142 anticancer drugs as output⁴³. LOBICO processes binary mutation data for 60 cancer genes as input features, including point mutations, insertions/deletions, amplifications, and gene fusions. The model outputs continuous IC50 values that are then binarized into sensitive/resistant classifications while retaining the continuous information through sample-specific weights. Multi-predictor models incorporating logic

combinations of mutations significantly outperformed single-gene predictors for 85% of drugs. One of the key advantages of LOBICO is its ability to retain continuous output information, which allows for robust and precise logic modeling. The algorithm also offers adjustable parameters and settings, enabling customization to achieve desired sensitivity and specificity levels, making it particularly suitable for clinical applications⁴³. LOBICO has demonstrated significant results in practical applications. For instance, in colorectal cancer, the algorithm identified Phosphatase and Tensin Homolog gene mutations as associated with sensitivity to Protein Kinase B inhibitors. Similarly, it revealed that Kirsten Rat Sarcoma Virus allele-specific mutations in colorectal cancer correlated with sensitivity to bicalutamide⁴³. These findings provide valuable insights for developing personalized treatment plans. However, it's important to note that the LOBICO approach faces certain limitations. The tissue specificity among different cancer types and the complexity of the model are recognized challenges⁴⁴. Future research should focus on expanding the model's applicability to different cancer types and optimizing the algorithm to enhance its interpretability and practicality.

MKL. The MKL model is a method that employs kernel regression techniques to integrate multiple data types, such as gene expression and genomic data, to predict drug sensitivity in cancer cell lines^{5,45}. MKL integrates different similarity measures expressed through various kernel functions and information from multiple sources or representations or feature subsets⁴⁵. Multiple Kernel Learning methods effectively handle different types of inputs and similarity measures from multiple sources or modalities. The combination of these different inputs can be achieved through three main approaches: linear combinations (using simple unweighted or weighted sums of kernels), nonlinear combinations (using multiplication or other nonlinear functions), or data-dependent combinations that adapt to local data distributions. For model evaluation and selection, the data is typically split into separate training and validation sets, with cross-validation procedures used to choose the best-performing kernel combination. This approach allows the model to leverage multiple notions of similarity rather than relying on a single kernel function, similar to the benefits seen in combining different classifiers. The merits of the MKL model encompass the integration of diverse data types, the utilization of non-linear modeling techniques, and the exploitation of multi-task learning to enhance prediction accuracy and generalization capability⁴⁶. By integrating biological data from multiple levels, including gene expression, genomic variations, and epigenetic modifications, the MKL model can provide a comprehensive characterization of the molecular features of cancer and the mechanisms underlying drug action⁴⁷. Moreover, the MKL model employs kernel functions to transform data into a high-dimensional feature space, allowing it to capture non-linear relationships within the data and improve modeling flexibility. Furthermore, the multi-task learning framework enables the MKL model to share information across various drugs and cell lines, enhancing the robustness and generalization capability of the predictions⁴⁸. Nevertheless, the MKL model also encounters several challenges. Firstly, generating multi-omics data necessitates substantial experimental costs and large sample sizes, which restricts the application scope of the MKL model. Secondly, integrating multiple data types elevates the complexity of modeling, requiring meticulous design of feature representation and fusion strategies. Furthermore, evaluating and validating the performance of the MKL model proves challenging, necessitating rigorous assessment on independent test sets and considering the influence of sample heterogeneity and batch effects. Despite these limitations, the MKL model presents a promising approach for drug sensitivity prediction by integrating multiple data types. Future research should focus on optimizing data collection strategies, such as employing high-throughput sequencing technologies and standardized experimental protocols, to minimize data generation costs. Simultaneously, improving modeling techniques, such as incorporating attention mechanisms and graph neural networks, can enhance the MKL model's capability to process

high-dimensional data and capture complex relationships. Finally, validating the performance of the MKL model on larger-scale independent test sets and comparing it with other cutting-edge methods will facilitate the evaluation of its potential for application in clinical decision support systems.

Integration of drug structure information or systems pharmacology with multi-omics data

DrugComboExplorer. DrugComboExplorer is a systems biology tool that integrates pharmacogenomic and multi-omic data to identify critical signaling pathways and predict synergistic drug combinations¹⁵. DrugComboExplorer integrates pharmacogenomics profiles of drugs, DNA-seq, gene copy number, DNA methylation, and RNA-seq data. The integration process follows a stepwise approach: first identifying driver networks from frequently mutated and copy number amplified genes, then combining these networks with co-expression and gene regulation networks. The model evaluates drug combinations by quantifying both collaborative targeted effects on the same driver signaling networks and complementary targeted effects on different network modules. It integrates various algorithms to generate driving signaling networks and utilizes systems pharmacology approaches to infer the efficacy and synergistic mechanisms of drug combinations. In studies of diffuse large B-cell lymphoma and prostate cancer, this tool enables researchers to predict and validate potential synergistic drug combinations using genomic data from cancer patients. It also assesses their effectiveness in targeting specific cancer signaling networks, which could potentially enhance treatment efficacy¹⁵. By integrating multi-omic data, DrugComboExplorer effectively predicts synergistic drug combinations for cancer, demonstrating the potential application of network-based drug efficacy screening methods in personalized cancer therapy and offering novel research directions and strategies for cancer treatment. Despite DrugComboExplorer's excellent performance in validation experiments, it has several limitations, such as the reliance on interactome data from public databases that may lack specificity, the necessity for validating predictive effects in additional cancer types, and the requirement for clinical trials to confirm the clinical efficacy of predicted combinations.

OncosynergyX. OncosynergyX is a computational predictive model that integrates multi-omic features of drugs and cell lines, such as gene expression, mutations, copy number variations, drug chemical structures, and protein targets, to predict the synergistic effects of anticancer drug combinations⁴⁹. The model primarily employs machine learning algorithms, including XGBoost and random forest, and optimizes model parameters through cross-validation. OncosynergyX integrates comprehensive pharmacological data with molecular information to train machine-learning models for predicting drug synergies. The key data source came from the DREAM AstraZeneca-Sanger Drug Combination Prediction Challenge, which provided one of the largest combinatorial cell line screening datasets including molecular, chemical, and biological data⁴⁹. The model processes multi-omics data through a biologically informed abridged feature set of 2121 features, condensed from the complete feature set of 111,168 features. The input includes monotherapy features, genomic context including expression, copy number variations, and mutations, and additional features like drug synergy networks. For data splitting, the study uses tenfold cross-validation, where some drug combinations in the test set contain drugs present in training data, as the model relies on both molecular structure and multi-omics cell line features for predictions. A key advantage of this method is its data-driven approach, which uses multiple types of biological data. Additionally, the model systematically evaluates various machine learning algorithms, identifying the XGBoost model as the most effective. The model's performance is further enhanced through hyperparameter tuning⁴⁹. However, the OncosynergyX model has certain limitations, including the wide range of drug concentrations covered by the output synergy scores, the potential for false positives in computational predictions, and the

inherent limitations of using IC50 as a measure of drug sensitivity. Notwithstanding these limitations, OncosynergyX exhibits superior performance and offers profound biological insights into the prediction of drug synergistic effects, thereby providing a valuable framework to guide experimental work in this field. Prospective validation studies are necessary to further confirm the translational potential of OncosynergyX in enhancing cancer treatment strategies.

Signaling pathways and signaling networks in drug interaction prediction

Random Walk with Restart (RWR). Random Walk with Restart (RWR) is a network-based computational method that predicts disease-associated genes by simulating the propagation of information through biological interaction networks⁵⁰. The method first constructs a multiplex network by integrating diverse molecular interaction data, including protein-protein interactions, pathway information, and gene co-expression patterns, and then employs RWR to calculate the diffusion distance from known disease genes to network nodes⁵⁰. The advantage of RWR lies in its ability to capture both direct and indirect functional relationships in the molecular network without requiring detailed mechanistic knowledge of disease pathways or extensive experimental validation data, thereby overcoming the limitations of traditional methods that focus solely on direct protein interactions or pathway annotations. Moreover, the multiplex network approach based on the integration of diverse interaction types facilitates the identification of disease genes with complementary functional roles. However, the RWR method also has limitations, such as the empirical nature of parameter selection during the construction of multiplex networks and the dependence on the quality and completeness of interaction databases. Despite these limitations, the method has demonstrated its potential from theory to application in various diseases, providing new insights for disease gene discovery and functional characterization of disease mechanisms.

Mathematical model based on the Epidermal Growth Factor Receptor (EGFR) signaling network.

This mathematical model, which is based on the current understanding of the EGFR signaling network, utilizes ordinary differential equations to describe the dynamic behavior of the network under various inhibition scenarios⁵¹. The study demonstrates how network topology influences drug interaction predictions by analyzing different combinations of target points. Each test simulation represents a unique intervention scenario, where inhibitors target either single nodes, parallel pathways, or serially-connected nodes in the signaling cascade. The validation approach focuses on comparing the relative effectiveness of different targeting strategies, particularly examining how upstream interventions affect downstream signal propagation. This systematic testing reveals that targeting multiple nodes in series produces stronger signal attenuation than targeting parallel pathways, providing insights for designing combination therapies based on network structure. The model's advantages include demonstrating that simultaneous inhibition of multiple upstream processes is more effective than individual inhibition, revealing the superadditive synergistic effects of cascading inhibition, and enabling the prediction of therapeutic outcomes for multi-kinase inhibitors. However, the model also has limitations, such as only considering a simplified subset of the EGFR network, omitting some key biological features, and being heavily dependent on the selection of initial conditions and parameters. Nevertheless, this study provides a valuable exploration of the effects and mechanisms of multi-target combination therapy, establishing the theoretical foundation for developing more refined and realistic signaling network models.

Three-node Enzymatic Network Model. The Three-node Enzymatic Network Model is a computational approach for studying and predicting drug combination effects⁵². This model consists of three enzymes, each existing in active or inactive states, interacting according to prescribed network structures. Using Michaelis-Menten kinetics and incorporating

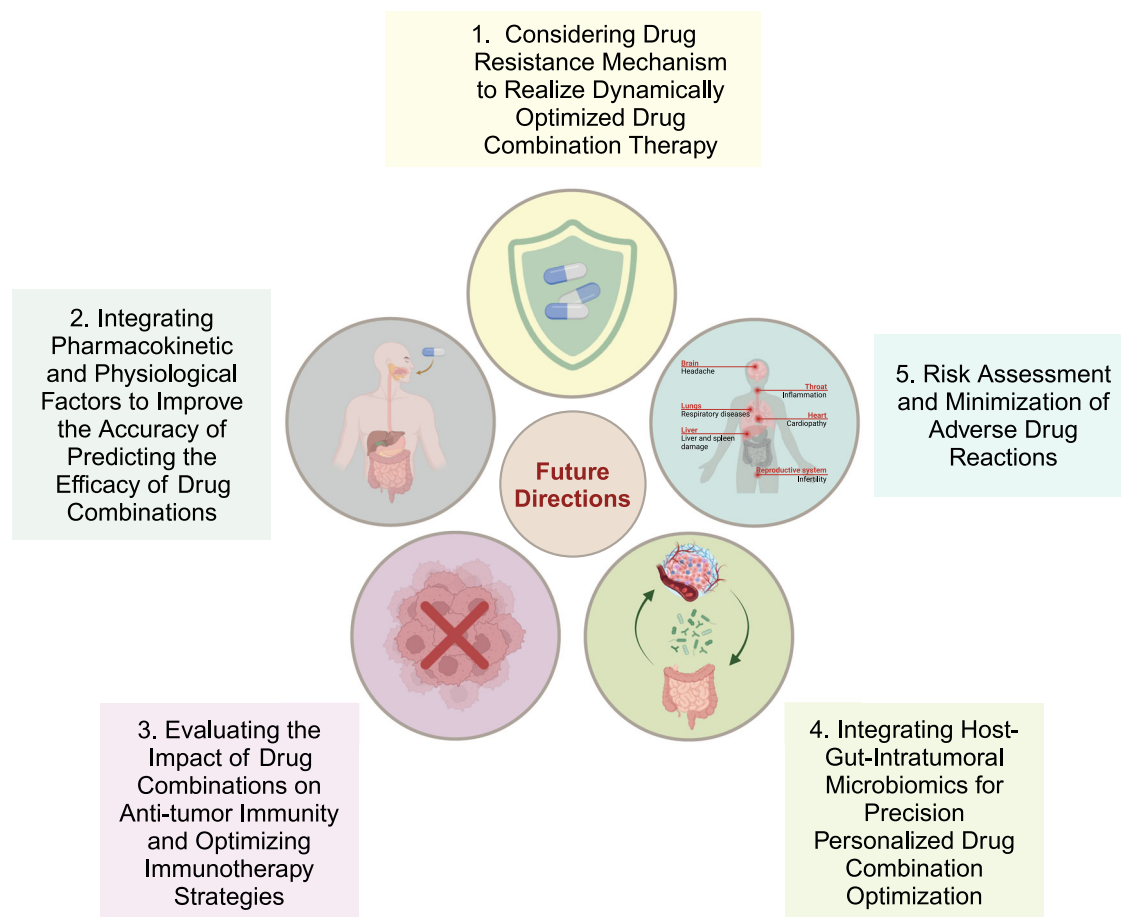


Fig. 4 | Future directions in algorithms for predicting drug synergy and antagonism based on multi-omics data. Five key future directions in drug combination therapy research, highlighting the integration of drug resistance mechanisms, pharmacokinetics, immunotherapy optimization, microbiome analysis, and adverse

reaction management. Common side effects are illustrated for different organ systems. This figure was created based on the tools provided by Biorender.com (accessed on 4/9/2024).

background regulations, the model explores all possible network topologies to identify consistent synergistic or antagonistic drug combination motifs. The Three-node Enzymatic Network model uses real-valued parameters representing enzyme kinetics. The input consists of network topology matrices defining node connections and their regulatory relationships. The model processes these inputs through ordinary differential equations based on Michaelis-Menten kinetics. The output is a real-valued combination index indicating drug interaction patterns. Unlike traditional machine learning approaches requiring training/test splits, this mechanistic model validates predictions through parameter space sampling, demonstrating that drug interaction patterns depend primarily on network structure rather than specific parameter values. Its key advantages include comprehensive exploration of network topologies, robustness to parameter variations, and the ability to represent simplifications of complex networks or closely connected target networks. The model's predictive power lies in its capacity to identify drug combination motifs based on network structure. However, it has limitations, such as being a significant simplification of real biological systems and focusing primarily on enzymatic interactions. Practical applications include rational drug combination design and elucidating the relationship between network topology and drug combination effects. Despite its potential, the model faces challenges in experimental validation, scaling to larger networks, integrating additional biological data, and translating theoretical predictions to clinical applications. The computational intensity required for analyzing all possible network topologies with multiple parameter sets is also a consideration. Nevertheless, the Three-node Enzymatic Network Model provides valuable insights into the

topological basis of drug synergy and antagonism, offering a promising approach for rational drug combination design in enzymatic systems. Future work may focus on addressing these challenges to enhance its applicability in more complex biological contexts and clinical settings.

Discussion

Consider drug resistance mechanism to realize dynamically optimized drug combination therapy

Various algorithms have been developed to predict drug synergy and antagonism. However, these methods face significant challenges in addressing drug resistance issues. In cancer, tumor cells exhibit high genetic instability and heterogeneity. During treatment, tumor cells may undergo evolution and adaptation, leading to changes in drug sensitivity and the development of drug resistance^{53–55}.

To address this issue, long-term multi-omics monitoring of patients susceptible to drug resistance is required (Fig. 4). By collecting comprehensive data from patients' genomes, transcriptomes, and proteomes throughout the treatment process and integrating genomic and transcriptomic data, we can dynamically monitor the molecular characteristics of tumors. Adjustments to the treatment regimen can be made in response to changes in drug resistance mechanisms, providing a deeper understanding of cancer progression mechanisms, uncovering new drug targets and biomarkers, and supporting precision medicine⁵³. This dynamic and integrated multi-omics analysis approach enhances our understanding of cancer mechanisms, aids in the discovery of new drug targets and biomarkers, and bolsters precision medicine efforts⁵⁶. For instance, using single-cell sequencing technology allows for high-

resolution gene expression and genomic analysis of tumor samples at various treatment stages to track the development and evolution of drug-resistant clones⁵⁷. Moreover, incorporating imaging histology data such as Computed Tomography, Magnetic Resonance Imaging, and Positron Emission Tomography scans provides information on cancer phenotype and tumor microenvironment, assisting in the assessment of drug effectiveness and the emergence of drug resistance^{58–60}. By integrating and analyzing these multidimensional data, predictive models can be constructed to real-time assess a patient's response to the current drug regimen and allow timely adjustments to the treatment strategy if the tumor exhibits signs of evolution or resistance. This personalized and dynamic approach to adjusting drug regimens aims to surpass the limitations of traditional static prediction methods, enhance drug effectiveness, and extend patient survival.

However, achieving this goal faces several challenges, including tumor heterogeneity, difficulties in patient recruitment for rare mutations, and low rates of genomics-guided therapy implementation. Context-dependent effects of targeted therapies, challenges in identifying reliable biomarkers, and limitations of genomics-only approaches further complicate precision oncology. These issues highlight the need for more comprehensive strategies integrating multiple data types and considering tumor evolution⁶¹. Future research will need to refine algorithms for multi-omics data analysis, develop more accurate and robust predictive models, and demonstrate their effectiveness and feasibility through prospective clinical trials. Additionally, enhancing multidisciplinary collaboration and integrating basic research, computational biology, and clinical practice will be crucial for progress in this field.

Integrating pharmacokinetic and physiological factors to improve the accuracy of predicting the efficacy of drug combinations

Predicting the synergistic and antagonistic effects of drug combinations using multi-omics data has emerged as a crucial strategy in precision medicine, exhibiting high accuracy and promising potential for clinical application. However, these predictions are predominantly performed in *in vitro* cell lines or animal models, whereas the process of drug action in the human body is significantly more intricate^{62–64}. Specifically, variations in the pharmacokinetic properties of different drugs *in vivo* can result in alterations in their concentration and exposure time within target organs, consequently influencing the actual efficacy of drug combinations⁶⁵. In particular, after oral administration, drugs need to undergo processes such as gastrointestinal absorption, hepatic first-pass effect, and systemic distribution before reaching the target organs to exert their effects^{66,67}. Differences in the physicochemical properties, dosage forms, and routes of administration of various drugs can all potentially impact their absorption and distribution *in vivo*⁶⁸. Furthermore, drugs in the body may also be subject to the effects of metabolic enzymes and transporters, leading to changes in their concentration and clearance rate⁶⁹. These factors may make it challenging to maintain the predicted drug combination ratios from *in vitro* studies *in vivo*, thus affecting the manifestation of synergistic or antagonistic effects. Moreover, the complexity of the tumor microenvironment may also influence the accumulation and action of drugs in target organs⁷⁰. Tumor tissues often exhibit characteristics such as abnormal angiogenesis, interstitial barriers, and acidic microenvironments, making it difficult for drugs to effectively penetrate and distribute⁷⁰. The concentration and distribution of different drugs in tumor tissues may be influenced by these factors, consequently altering the efficacy of drug combinations.

Although some studies have attempted to predict drug distribution and target organ concentrations *in vivo* using pharmacokinetic models and physiological modeling, these methods often rely on animal experimental data and limited human trial data, making it difficult to fully reflect individual differences and tumor heterogeneity⁷¹. Therefore, how to integrate pharmacokinetic and pharmacodynamic factors into multi-omics prediction models to more accurately predict the

efficacy of drug combinations in individual patients remains a pressing issue to be addressed (Fig. 4). Future research directions could include developing more sophisticated physiological modeling techniques that account for the distribution and clearance processes of drugs across various organs and tumor tissues, as well as integrating multi-omics data with clinical pharmacokinetic data to construct individualized drug concentration-time curve prediction models. Through these efforts, it is hoped that more precise and personalized optimization of drug combinations can be achieved, enhancing the success rate of precision cancer therapy.

Evaluating the impact of drug combinations on anti-tumor immunity and optimizing immunotherapy strategies

While drug combinations may demonstrate promising synergistic effects *in vitro*, they often face more complex biological challenges when administered *in vivo*. Among these challenges, the potential negative impact of drug combinations on immune cells is a critical factor that demands attention⁷². For instance, combining therapies can increase toxicity, particularly immune-related adverse effects⁷². Moreover, some targeted therapies, such as checkpoint inhibitors, can activate T cells but may also trigger autoimmune-related adverse reactions, thereby indirectly affecting anti-tumor immunity. Immune checkpoint inhibitors bolster anti-tumor immunity by unleashing T cells, but they can precipitate immune-related adverse events that mimic autoimmune conditions. These immune-related adverse events manifest across diverse organ systems with a spectrum of severity. Notably, combination immunotherapy regimens may amplify both the frequency and intensity of these immune-mediated side effects⁷³. However, many current drug combination screening and prediction algorithms rely primarily on the direct response of tumor cells, such as proliferation inhibition and apoptosis induction, without adequately considering the impact of drugs on the immune microenvironment. This may lead to the selection of drug combinations with poor *in vivo* efficacy, as they may impair the body's anti-tumor immune function.

To address this issue, future drug combination screening and optimization strategies should consider the effects of drugs on the immune system more comprehensively (Fig. 4). Immunotoxicity and immunomodulatory effects can be incorporated into the evaluation criteria for drug combinations. The impact of drug combinations on anti-tumor immunity can be evaluated through methods such as *in vitro* co-culture of immune cells and tumor cells⁷⁴. Furthermore, by leveraging AI and machine learning techniques to integrate multi-omics data and immune phenotype data, the immunomodulatory effects and overall efficacy of drug combinations *in vivo* can be predicted more accurately. In conclusion, considering the impact of drug combinations on the anti-tumor immune microenvironment is crucial for optimizing combination therapies and enhancing clinical efficacy. Future research should focus on developing more comprehensive and precise drug combination optimization strategies that take into account both tumor cell and immune cell responses, in order to maximize the clinical benefit of combination therapies.

Integrating host-gut-intratatumoral microbiomics for precision personalized drug combination optimization

Drug combination therapy is a crucial strategy in precision oncology; however, current algorithms for optimizing drug combinations have not fully considered their impact on the gut and intratumoral microbial communities. Accumulating evidence suggests that microbial communities play a pivotal role in tumor initiation, progression, and treatment response⁷⁵.

First, gut microbiota can modulate the host's immune function and drug metabolism, thereby influencing drug efficacy and toxicity⁷⁶. Certain drug combinations may alter the composition and function of the gut microbial community, potentially influencing the antitumor effect⁷⁷. Furthermore, gut microbiota can influence the host's response to drugs through metabolic reprogramming⁷⁸. Second, the microbial community in the tumor microenvironment is also closely associated with drug

Table 4 | Multi-omics data integration for addressing key challenges in cancer treatment

Challenge	Relevant Omics Data	Representative Methods/Algorithms	Expected Outcome
Drug Resistance	Genomics; Transcriptomics; Proteomics	Deep learning networks	Dynamic prediction of resistance development
Pharmacokinetics/Dynamics	Genomics; Metabolomics; Proteomics	Physiologicalbased modeling Machine learning regression Network analysis	Improved drug distribution prediction
Immune Response	Immunogenomics; Transcriptomics; Proteomics	Immune cell deconvolution Multitask learning Systems biology modeling	Better immunotherapy response prediction
Microbiome Impact	Genomics; Metabolomics; Proteomics	Graph Convolutional Networks Deep Learning Methods Multi-Omics Predictive Models	Optimization of microbiomeaware therapy
Adverse Reactions	Genomics; Transcriptomics; Proteomics; Multi-omics Data Integration	Risk prediction models Bayesian networks Ensemble methods	Reduced adverse event occurrence

efficacy. Research has demonstrated that certain bacteria can selectively accumulate within tumors and promote tumor growth and drug tolerance through mechanisms such as inducing immunosuppression or producing specific metabolites⁷⁹. Conversely, some beneficial bacteria can enhance the efficacy of drugs, such as immune checkpoint inhibitors⁷⁷. Consequently, drug combinations may indirectly influence the therapeutic response of tumors by reshaping the intratumoral microbial community.

Current algorithms for drug combination screening and optimization are primarily based on the characteristics of tumor cells and host genomes and have not yet incorporated microbiome data. To more accurately predict the efficacy and toxicity of drug combinations, future algorithms should integrate host genomics, tumor genomics, and microbiome information to construct multi-omics predictive models (Fig. 4). Simultaneously, by employing machine learning and other techniques to identify key microbial biomarkers and metabolic pathways, we can guide the optimization of drug combinations and personalized therapies⁸⁰. In conclusion, considering the impact of drug combinations on gut and intratumoral microbial communities is crucial for improving drug efficacy and reducing toxicity. Integrating microbiomics into drug combination optimization algorithms holds promise for achieving more precise and personalized cancer treatment decisions.

Risk assessment and minimization of adverse drug reactions

Although combination therapy can improve the efficacy of tumor treatment, it may also increase the risk of adverse drug reactions (ADRs). ADRs are defined as harmful and unintended responses that occur at normal therapeutic doses of medication⁸¹. The adverse reaction profiles of different drugs may overlap or interact synergistically, potentially leading to an increased incidence or exacerbation of ADRs. Consequently, when designing combination therapy regimens, it is imperative to comprehensively assess the ADR risk associated with specific drug combinations (Fig. 4). The primary objective of combination drug therapy is to maximize antitumor efficacy while minimizing the occurrence and severity of ADRs. To attain this objective, it is crucial to continuously refine ADR prediction models and integrate efficacy and safety data for various drug combinations. By carefully balancing the benefits and risks of drugs, healthcare professionals can provide patients with the most appropriate personalized combination therapy regimen. This necessitates close collaboration among multi-disciplinary teams to highly integrate expertise from fields such as pharmacology, clinical medicine, and computer science, facilitating the continuous improvement of the risk assessment system for ADRs, ultimately benefiting a broad spectrum of cancer patients.

Table 4 presents a comprehensive framework for integrating multi-omics data to address five key challenges in cancer treatment: drug

resistance, pharmacokinetics/dynamics, immune response, microbiome impact, and adverse reactions. For each challenge, the table outlines the relevant omics data types required, representative computational methods and algorithms, and the expected outcomes of these integrated approaches. This systematic organization provides a roadmap for tackling complex cancer treatment issues through multi-dimensional data analysis. The table serves as a foundation for the detailed discussion that follows, where each challenge is thoroughly examined in terms of current limitations, potential solutions, and future directions. As explored in section “Consider Drug Resistance Mechanism to Realize Dynamically Optimized Drug Combination Therapy”, the first challenge of drug resistance requires dynamic monitoring through multiple omics layers to track tumor evolution and adapt treatment strategies accordingly.

Conclusion

This article presents a systematic review of drug synergy combination prediction models based on multi-omics data. With the advancement of high-throughput sequencing technologies, biological big data across various dimensions, including genomics, epigenomics, and transcriptomics, are experiencing explosive growth. Leveraging these vast heterogeneous data to develop drug synergy prediction models holds promise for expediting the discovery and optimization of synergistic drug combinations.

First, we elucidate the concepts of drug synergy and antagonism and highlight the pivotal role of AI algorithms in optimizing drug combinations. Subsequently, we summarize the commonalities among various prediction algorithms in terms of data input, feature extraction, feature selection, model validation, and other aspects. Furthermore, this article introduces multiple strategies and technical approaches for integrating multi-omics data, such as kernel regression, machine learning, graph networks, and metabolic network simulation. Moreover, we highlight several representative drug synergy prediction models across different dimensions, encompassing genomics, epigenomics, and transcriptomics. Models based on genomic data, such as DrugComboRanker and AuDNNsynergy, primarily utilize gene expression profiles and mutation data to construct synergy prediction frameworks. Models based on epigenomic data, such as MethylMix, focus on the regulatory mechanisms of epigenetic modifications in drug synergy. Models based on transcriptomic data, such as CDA and DIGRE, emphasize the analysis of drug synergy mechanisms from the perspective of transcriptional changes.

Despite certain progress, multi-omics-based drug synergy combination prediction still faces numerous challenges, including the standardization and integration of multi-source heterogeneous data, the generalizability and interpretability of models, and the experimental validation of prediction results. In the future, it is imperative to develop more universal and robust multi-omics data integration frameworks, enhance the interpretability of models, elucidate the underlying mechanisms of drug synergy, and

concurrently reinforce the experimental validation and clinical translation research of model prediction results.

In conclusion, drug synergy prediction based on multi-omics big data is a burgeoning research field. The innovation and refinement of models are expected to provide novel insights for optimizing synergistic drug combinations, foster the development of personalized precision medicine, and ultimately benefit human health. It is anticipated that with the increasing maturity of AI, multi-omics, and other technologies, drug synergy combination prediction models will undoubtedly embrace broader application prospects in the future.

Abbreviation

DREAM	Dialogue on Reverse Engineering Assessment and Methods
AI	artificial intelligence
MKL	multiple kernel learning
PPI	protein–protein interaction
BI	Bliss Independence
CI	Combination Index
CMAP	Connectivity Map
TCGA	The Cancer Genome Atlas
DNA	deoxyribonucleic acid
CDA	combinatorial drug assembler
DIGRE	drug-induced genomic residual effect
MR	master regulator
LOBICO	logic optimization for binary input to continuous output
RWR	random walk with restart
EGFR	epidermal growth factor receptor
ADR	adverse drug reaction
AUC	area under the curve
ROC	receiver operating characteristic

Received: 6 October 2024; Accepted: 3 January 2025;

Published online: 19 January 2025

References

- Jia, J. et al. Mechanisms of drug combinations: interaction and network perspectives. *Nat. Rev. Drug Discov.* **8**, 111–128 (2009).
- Chou, T.-C. Theoretical basis, experimental design, and computerized simulation of synergism and antagonism in drug combination studies. *Pharmacol. Rev.* **58**, 621–681 (2006).
- Collins, F. S. et al. New goals for the US human genome project: 1998–2003. *Science* **282**, 682–689 (1998).
- Hoheisel, J. D. Microarray technology: beyond transcript profiling and genotype analysis. *Nat. Rev. Genet.* **7**, 200–210 (2006).
- Costello, J. C. et al. A community effort to assess and improve drug sensitivity prediction algorithms. *Nat. Biotechnol.* **32**, 1202–1212 (2014).
- Preuer, K. et al. DeepSynergy: predicting anti-cancer drug synergy with Deep Learning. *Bioinformatics* **34**, 1538–1546 (2018).
- Huang, L. et al. DrugComboRanker: drug combination discovery based on target network analysis. *Bioinformatics* **30**, i228–i236 (2014).
- Bulusu, K. C. et al. Modelling of compound combination effects and applications to efficacy and toxicity: state-of-the-art, challenges and perspectives. *Drug Discov. Today* **21**, 225–238 (2016).
- Cheng, F. & Zhao, Z. Machine learning-based prediction of drug–drug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties. *J. Am. Med. Inf. Assoc.* **21**, e278–e286 (2014).
- Madani Tonekaboni, S. A., Soltan Ghorraie, L., Manem, V. S. K. & Haibe-Kains, B. Predictive approaches for drug combination discovery in cancer. *Brief Bioinform.* **19**, 263–276 (2018).
- Stokes, J. M. et al. A deep learning approach to antibiotic discovery. *Cell* **180**, 688–702.e613 (2020).
- Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
- Bansal, M. et al. A community computational challenge to predict the activity of pairs of compounds. *Nat. Biotechnol.* **32**, 1213–1222 (2014).
- Zhang, T., Zhang, L., Payne, P. R. & Li, F. Synergistic drug combination prediction by integrating multiomics data in deep learning models. *Translational Bioinformatics for Therapeutic Development*, 223–238 (2021).
- Huang, L. et al. Driver network as a biomarker: systematic integration and network modeling of multi-omics data to derive driver signaling pathways for drug combination prediction. *Bioinformatics* **35**, 3709–3717 (2019).
- Bliss, C. I. The toxicity of poisons applied jointly 1. *Ann. Appl. Biol.* **26**, 585–615 (1939).
- Chou, T.-C. & Talalay, P. Quantitative analysis of dose–effect relationships: the combined effects of multiple drugs or enzyme inhibitors. *Adv. Enzym. Regul.* **22**, 27–55 (1984).
- Zhou, J. et al. Deep learning sequence-based ab initio prediction of variant effects on expression and disease risk. *Nat. Genet.* **50**, 1171–1179 (2018).
- Zitnik, M., Agrawal, M. & Leskovec, J. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics* **34**, i457–i466 (2018).
- Bordbar, A. et al. Personalized whole-cell kinetic models of metabolism for discovery in genomics and pharmacodynamics. *Cell Syst.* **1**, 283–292 (2015).
- Guebila, M. B. & Thiele, I. Model-based dietary optimization for late-stage, levodopa-treated, Parkinson’s disease patients. *NPJ Syst. Biol. Appl.* **2**, 16013 (2016).
- Nilsson, A. & Nielsen, J. Genome scale metabolic modeling of cancer. *Metab. Eng.* **43**, 103–112 (2017).
- Lamb, J. et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* **313**, 1929–1935 (2006).
- Kuhn, M. et al. STITCH 3: zooming in on protein–chemical interactions. *Nucleic Acids Res.* **40**, D876–D880 (2012).
- Stark, C. et al. BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* **34**, D535–D539 (2006).
- Rigatti, S. J. Random forest. *J. Insur. Med.* **47**, 31–39 (2017).
- Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B: Stat. Methodol.* **67**, 301–320 (2005).
- Cedoz, P. L., Prunello, M., Brennan, K. & Gevaert, O. MethylMix 2.0: an R package for identifying DNA methylation genes. *Bioinformatics* **34**, 3044–3046 (2018).
- Lee, J. H. et al. CDA: combinatorial drug discovery using transcriptional response modules. *PLoS One* **7**, e42573 (2012).
- Li, X. et al. Prediction of synergistic anti-cancer drug combinations based on drug target network and drug induced gene expression profiles. *Artif. Intell. Med.* **83**, 35–43 (2017).
- Zhao, L., Au, J. L.-S. & Wientjes, M. G. Comparison of methods for evaluating drug–drug interaction. *Front. Biosci. (Elite Ed.)* **2**, 241 (2010).
- Sun, X., Vilar, S. & Tatonetti, N. P. High-throughput methods for combinatorial drug discovery. *Sci. Transl. Med.* **5**, 205rv201–205rv201 <https://doi.org/10.1126/scitranslmed.3006667> (2013).
- Li, P. et al. Large-scale exploration and analysis of drug combinations. *Bioinformatics* **31**, 2007–2016 (2015).
- Saez-Rodriguez, J. et al. Crowdsourcing biomedical research: leveraging communities as innovation engines. *Nat. Rev. Genet.* **17**, 470–486 (2016).
- Stolovitzky, G., Monroe, D. & Califano, A. Dialogue on reverse-engineering assessment and methods: the DREAM of high-throughput pathway inference. *Ann. N.Y. Acad. Sci.* **1115**, 1–22 (2007).
- Gerlee, P. et al. Searching for synergies: matrix algebraic approaches for efficient pair screening. *PLoS One* **8**, e68598 (2013).

37. Wang, X. et al. PRODeepSyn: predicting anticancer synergistic drug combinations by embedding cell lines with protein-protein interaction network. *Brief Bioinform.* **23**, <https://doi.org/10.1093/bib/bbab587> (2022).
38. Gan, Y., Huang, X., Guo, W., Yan, C. & Zou, G. Predicting synergistic anticancer drug combination based on low-rank global attention mechanism and bilinear predictor. *Bioinformatics* **39**, <https://doi.org/10.1093/bioinformatics/btad607> (2023).
39. Zheng, S. et al. DrugComb update: a more comprehensive drug sensitivity data repository and analysis portal. *Nucleic Acids Res.* **49**, W174–W184 (2021).
40. Gaulton, A. et al. The ChEMBL database in 2017. *Nucleic Acids Res.* **45**, D945–D954 (2017).
41. Luck, K. et al. A reference map of the human binary protein interactome. *Nature* **580**, 402–408 (2020).
42. Ghandi, M. et al. Next-generation characterization of the cancer cell line encyclopedia. *Nature* **569**, 503–508 (2019).
43. Knijnenburg, T. A. et al. Logic models to predict continuous outputs based on binary inputs with an application to personalized cancer therapy. *Sci. Rep.* **6**, 36812 (2016).
44. Beijersbergen, R. L., Wessels, L. F. & Bernards, R. Synthetic lethality in cancer therapeutics. *Annu. Rev. Cancer Biol.* **1**, 141–161 (2017).
45. Gönen, M. & Alpaydin, E. Multiple kernel learning algorithms. *J. Mach. Learn. Res.* **12**, 2211–2268 (2011).
46. Ali, M. & Aittokallio, T. Machine learning and feature selection for drug response prediction in precision oncology applications. *Biophys. Rev.* **11**, 31–39 (2019).
47. Chiu, Y. C. et al. Predicting drug response of tumors from integrated genomic profiles by deep neural networks. *BMC Med. Genom.* **12**, 18 (2019).
48. Yuan, H., Paskov, I., Paskov, H., Gonzalez, A. J. & Leslie, C. S. Multitask learning improves prediction of cancer drug sensitivity. *Sci. Rep.* **6**, 31619 (2016).
49. Celebi, R., Bear Don't Walk, O. T., Movva, R., Alpsy, S. & Dumontier, M. In-silico prediction of synergistic anti-cancer drug combinations using multi-omics data. *Sci. Rep.* **9**, 8949 (2019).
50. Valdeolivas, A. et al. Random walk with restart on multiplex and heterogeneous biological networks. *Bioinformatics* **35**, 497–505 (2018).
51. Araujo, R. P., Petricoin, E. F. & Liotta, L. A. A mathematical model of combination therapy using the EGFR signaling network. *Biosystems* **80**, 57–69 (2005).
52. Yin, N. et al. Synergistic and antagonistic drug combinations depend on network topology. *PLoS One* **9**, e93960 (2014).
53. Dagogo-Jack, I. & Shaw, A. T. Tumour heterogeneity and resistance to cancer therapies. *Nat. Rev. Clin. Oncol.* **15**, 81–94 (2018).
54. Greaves, M. & Maley, C. C. Clonal evolution in cancer. *Nature* **481**, 306–313 (2012).
55. Burrell, R. A. & Swanton, C. Tumour heterogeneity and the evolution of polyclonal drug resistance. *Mol. Oncol.* **8**, 1095–1111 (2014).
56. Chakraborty, S., Hosen, M. I., Ahmed, M. & Shekhar, H. U. Onco-multi-OMICS approach: a new frontier in cancer research. *Biomed. Res. Int.* **2018**, 9836256 (2018).
57. Maynard, A. et al. Therapy-induced evolution of human lung cancer revealed by single-cell RNA sequencing. *Cell* **182**, 1232–1251 e1222 (2020).
58. Aerts, H. J. et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat. Commun.* **5**, 4006 (2014).
59. Gillies, R. J., Kinahan, P. E. & Hricak, H. Radiomics: images are more than pictures, they are data. *Radiology* **278**, 563–577 (2016).
60. Lambin, P. et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat. Rev. Clin. Oncol.* **14**, 749–762 (2017).
61. Senft, D., Leiserson, M. D. M., Ruppin, E. & Ronai, Z. A. Precision oncology: the road ahead. *Trends Mol. Med.* **23**, 874–898 (2017).
62. Benet, L. Z. & Zia-Amirhosseini, P. Basic principles of pharmacokinetics. *Toxicol. Pathol.* **23**, 115–123 (1995).
63. Kola, I. & Landis, J. Can the pharmaceutical industry reduce attrition rates? *Nat. Rev. Drug Discov.* **3**, 711–716 (2004).
64. Mak, I. W., Evaniew, N. & Ghert, M. Lost in translation: animal models and clinical trials in cancer treatment. *Am. J. Transl. Res.* **6**, 114 (2014).
65. Bohnert, T. & Gan, L. S. Plasma protein binding: from discovery to development. *J. Pharm. Sci.* **102**, 2953–2994 (2013).
66. Di, L. et al. Evidence-based approach to assess passive diffusion and carrier-mediated drug transport. *Drug Discov. Today* **17**, 905–912 (2012).
67. Sahinovic, M. M., Struys, M. M. & Absalom, A. R. Clinical pharmacokinetics and pharmacodynamics of propofol. *Clin. Pharmacokinet.* **57**, 1539–1558 (2018).
68. Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **64**, 4–17 (2012).
69. International Transporter, C. et al. Membrane transporters in drug development. *Nat Rev Drug Discov* **9**, 215–236 (2010).
70. Minchinton, A. I. & Tannock, I. F. Drug penetration in solid tumours. *Nat. Rev. Cancer* **6**, 583–592 (2006).
71. Rowland, M., Peck, C. & Tucker, G. Physiologically-based pharmacokinetics in drug development and regulatory science. *Annu. Rev. Pharmacol. Toxicol.* **51**, 45–73 (2011).
72. Mahoney, K. M., Rennert, P. D. & Freeman, G. J. Combination cancer immunotherapy and new immunomodulatory targets. *Nat. Rev. Drug Discov.* **14**, 561–584 (2015).
73. Day, D. & Hansen, A. R. Immune-related adverse events associated with immune checkpoint inhibitors. *BioDrugs* **30**, 571–584 (2016).
74. Courau, T. et al. Cocultures of human colorectal tumor spheroids with immune cells reveal the therapeutic potential of MICA/B and NKG2A targeting for cancer treatment. *J. Immunother. Cancer* **7**, 1–14 (2019).
75. Heshiki, Y. et al. Predictable modulation of cancer treatment outcomes by the gut microbiota. *Microbiome* **8**, 28 (2020).
76. Alexander, J. L. et al. Gut microbiota modulation of chemotherapy efficacy and toxicity. *Nat. Rev. Gastroenterol. Hepatol.* **14**, 356–365 (2017).
77. Gopalakrishnan, V. et al. Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. *Science* **359**, 97–103 (2018).
78. Zimmermann, M., Zimmermann-Kogadeeva, M., Wegmann, R. & Goodman, A. L. Mapping human microbiome drug metabolism by gut bacteria and their genes. *Nature* **570**, 462–467 (2019).
79. Geller, L. T. et al. Potential role of intratumor bacteria in mediating tumor resistance to the chemotherapeutic drug gemcitabine. *Science* **357**, 1156–1160 (2017).
80. Pasolli, E., Truong, D. T., Malik, F., Waldron, L. & Segata, N. Machine learning meta-analysis of large metagenomic datasets: tools and biological insights. *PLoS Comput. Biol.* **12**, e1004977 (2016).
81. Alomar, M. J. Factors affecting the development of adverse drug reactions (Review article). *Saudi Pharm. J.* **22**, 83–94 (2014).

Acknowledgements

This study was partly supported by the National Natural Science Foundation of China (32322087 to S.Y.), HKU Seed Fund for Translational and Applied Research (2203100624 to S.Y.), the Natural Science Foundation of Guangdong Province (2021A1515012593), the National Natural Science Foundation of China (82373129, 82260546, 82172811, 82172750, 81802257), the Guangdong Basic and Applied Basic Research Foundation (Guangdong–Guangzhou Joint Funds) (2022A1515111212), and the Science and Technology Program of Guangzhou (2023A04J1257). The funding sources had no role in the study design, data collection, analysis, interpretation, or writing of the report.

Author contributions

P.L., S.Y., and Q.C. contributed to conception and manuscript design. J.C., A.L., A.J., and C.Q. drafted the manuscript and prepared the figures. J.C., A.L., and Z.L. collected the related references. J.C., A.L., A.J., C.Q., Z.L., Q.C., S.Y., and P.L. participated in the revision of the manuscript. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Quan Cheng, Shuofeng Yuan or Peng Luo.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025