



# Bioinformatics Research on Drug Sensitivity Prediction

Yaojia Chen<sup>1</sup>, Liran Juan<sup>2</sup>, Xiao Lv<sup>3\*</sup> and Lei Shi<sup>4\*</sup>

<sup>1</sup>Yangtze Delta Region Institute (Quzhou), University of Electronic Science and Technology of China, Quzhou, China, <sup>2</sup>School of Life Science and Technology, Harbin Institute of Technology, Harbin, China, <sup>3</sup>Beidahuang Industry Group General Hospital, Harbin, China, <sup>4</sup>Department of Spine Surgery Changzheng Hospital, Naval Medical University, Shanghai, China

Modeling-based anti-cancer drug sensitivity prediction has been extensively studied in recent years. While most drug sensitivity prediction models only use gene expression data, the remarkable impacts of gene mutation, methylation, and copy number variation on drug sensitivity are neglected. Drug sensitivity prediction can both help protect patients from some adverse drug reactions and improve the efficacy of treatment. Genomics data are extremely useful for drug sensitivity prediction task. This article reviews the role of drug sensitivity prediction, describes a variety of methods for predicting drug sensitivity. Moreover, the research significance of drug sensitivity prediction, as well as existing problems are well discussed.

**Keywords:** drug sensitivity, database, machine learning, deep learning, anti-cancer

## OPEN ACCESS

### Edited by:

Xiujuan Lei,  
Shaanxi Normal University, China

### Reviewed by:

Jiangning Song,  
Monash University, Australia  
Liang Yu,  
Xidian University, China

### \*Correspondence:

Xiao Lv  
xiaolvxl@163.com  
Lei Shi  
slspine@163.com

### Specialty section:

This article was submitted to  
Experimental Pharmacology and Drug  
Discovery,  
a section of the journal  
Frontiers in Pharmacology

**Received:** 22 October 2021

**Accepted:** 18 November 2021

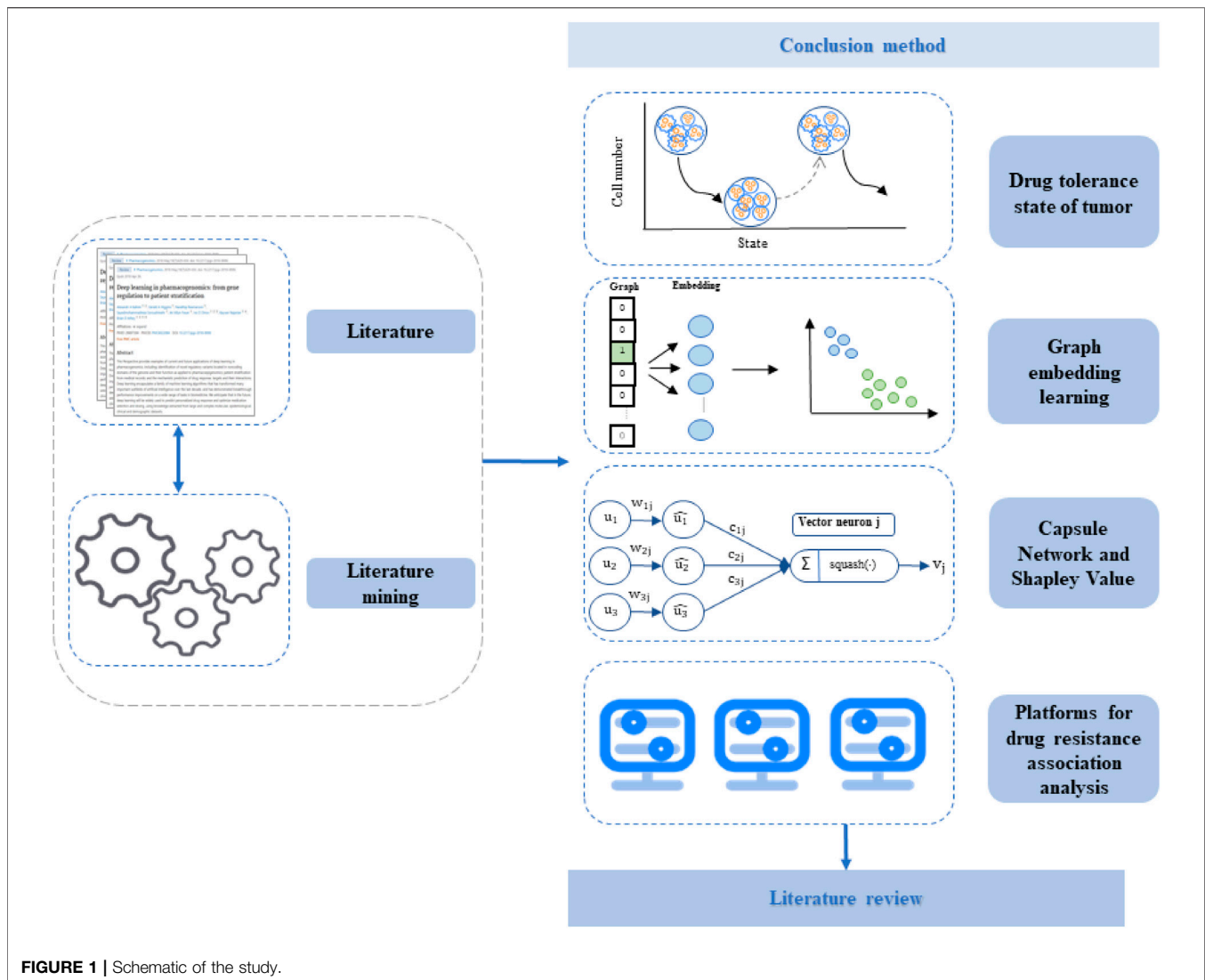
**Published:** 09 December 2021

### Citation:

Chen Y, Juan L, Lv X and Shi L (2021)  
Bioinformatics Research on Drug  
Sensitivity Prediction.  
Front. Pharmacol. 12:799712.  
doi: 10.3389/fphar.2021.799712

## 1 INTRODUCTION

With the significant technological advancements, a variety of modalities have been developed to predict the sensitivity of tumors to anti-cancer drugs, which can improve drug efficacy and reduce adverse effects and the financial burden of treatment on patient. The sensitivity of tumors to anti-cancer drugs can be assessed by using patient cell lines, which can facilitate the use of synergistic regimens (Liu et al., 2016). Unfortunately, this process requires substantial time and carries a high risk (Hanna, 2006; Russo, 2015; Cheng et al., 2019; Zhuang et al., 2020). Moreover, tumor drug resistance is another critical problem in the research and development of anti-cancer drugs and the medical field (Liu et al., 2020; Zhang et al., 2020a). It is well known that traditional cancer treatment approaches mainly aim to eradicate rapidly proliferating tumor cells (Restifo et al., 2016; Cheng et al., 2018; O'Donnell et al., 2018; Liu and Chen, 2020; Liu et al., 2021a). However, existing evidence illustrates that tumor cell subgroups can survive by resisting treatment through resistance mechanism, and these cells will finally evolve into drug-resistant tumor cells. It is challenging to elucidate the mechanisms by which tumors acquire drug resistance, predict the evolution of drug-resistant tumors, and determine appropriate strategies to eliminate recalcitrant cells. In addition, the identification of mutations that increase sensitivity to anti-cancer drugs and formulation of appropriate treatment plans for patient groups with specific genomic mutations have essential roles in the development of targeted therapies and achievement of precision treatment for human cancer. However, the traditional strategy for predicting drug sensitivity based on the similarity with known mutations has limitations (Carr et al., 2016; Jennifer et al., 2016; Schmitt et al., 2016; Li et al., 2017; Song et al., 2020; Qi et al., 2021). Meanwhile, the large amount of data resources related to markers of anti-cancer drug sensitivity need to be integrated. Most importantly, among the cancers, some are transmittable, thus triggering panic. Therefore, therapeutic drugs are urgently needed that can stop cancer transmission. The transmission of drug-resistant strains is an extremely serious major public health issue. Meanwhile, drug-



**FIGURE 1** | Schematic of the study.

resistant mutant strains of HIV-1 have hampered anti-viral treatment. It is necessary to study the mutation and subtype characteristics of HIV-1 recombinants, evolutionary principles, and drug resistance to develop vaccines and related drugs and implement preventative and control measures for AIDS (Castro-Nallar et al., 2012; Hemelaar, 2012; Shaw and Hunter, 2012).

This review focuses on several strategies related to drug sensitivity analysis. We first discuss the types of evolutionary models of drug resistance studies on drug-tolerant tumor cells after treatment. Meanwhile, we describe the process of models, including the construction of a time-series biological network and the evolution prediction of the tumor drug resistance state *via* k-means++ clustering, random walk, and other machine learning methods (Oxnard and Geoffrey, 2016; Hangauer et al., 2017; Recasens and Munoz, 2019; Yu et al., 2020a; Cheng et al., 2021). Second, we describe the strategy of sensitivity prediction of tumors to anti-cancer drugs using graph representation learning. This strategy can explain the

mechanism by which cancer develops, and most importantly provide reliable evidence for cancer treatment to promote the development of bioinformatics. Third, we discuss the strategy of developing a collaborative drug sensitivity analysis platform that can provide specific cancer cell lines with optimal stimulatory or inhibitory candidate drug molecules. This strategy provides new technical solutions for the development of anti-cancer drugs and overcomes the insufficiency of deep learning modeling methods for analyzing anti-cancer drug sensitivity (Jaiswal et al., 2018; Zhao et al., 2018; Azad et al., 2019; Cheng et al., 2020; An and Yu, 2021; Shang et al., 2021). The fourth strategy mainly aims to establish platforms for drug resistance association analysis to reduce the blind use of drug-resistant HIV strains and improve the effectiveness of AIDS treatment. There is an urgent need to develop drugs for both AIDS and cancer to limit deleterious effects in patients. Therefore, it is necessary to review bioinformatics research into drug sensitivity prediction. The outline of the essay is provided in **Figure 1**.

## 2 METHODS OF RESEARCH ON DRUG SENSITIVITY

### 2.1 Evolutionary Model Based on Drug Resistance

#### 2.1.1 Introduction to Methods

This type of researches mainly aimed to explore the drug resistance mechanisms of various anti-cancer drugs and the evolutionary direction of the drug resistance state. They target the characteristics of drug-resistant tumors during treatment through machine learning and deep learning methods, as well as the use of high-throughput pharmaceutical informatics data. Scientists generally conduct research from four aspects. 1) The first aspect is mainly concerned with the analysis of different drug resistance mechanisms arising from pan-cancer and tumor drug resistance based on the different tumor cell lines after treatment. 2) The second is mainly focused on the construction of a prediction model for the evolution of tumor drug resistance. In this part, gene mutations in various cell lines can be added to the prediction model according to the mutation frequency of the gene as the evolutionary condition of resistance. 3) The third aspect is mainly concerned with the design of drug application strategies that interfere with the tumor drug tolerance state. This means that the treatment plan can be devised according to genes affected by existing drugs and the classification of anti-tumor action principles. 4) The fourth aspect of methods mainly focus on the verification of the predicted medication plan through gene chip and cell experiments. Finally, the relevant interference drug plan can be developed using the prediction model, and then gene chip detection can be used for comparisons with the results obtained from the prediction model to verify the accuracy of the model.

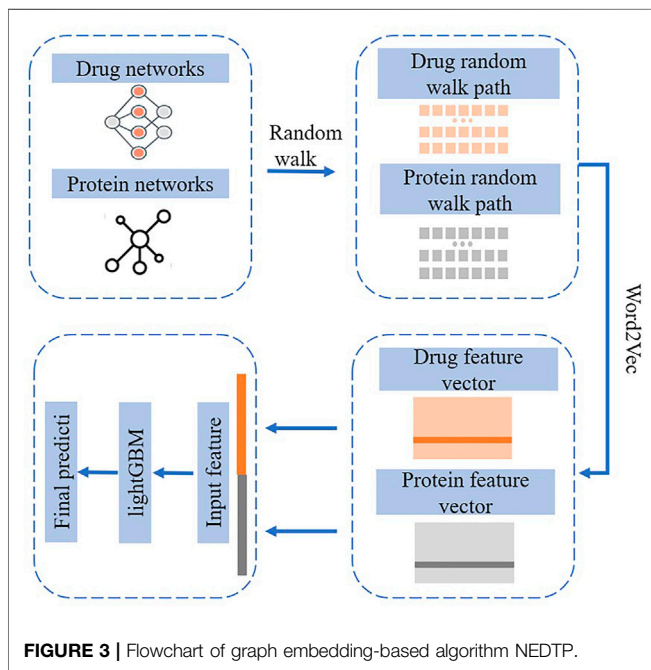
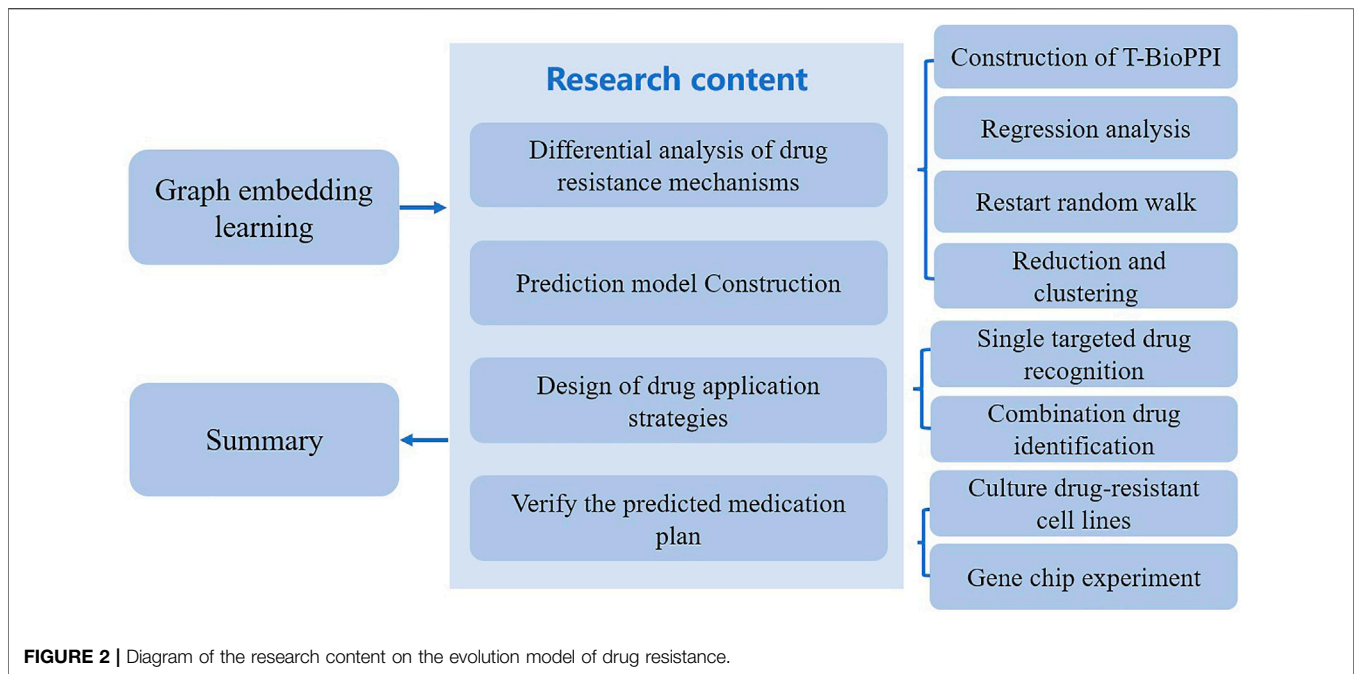
#### 2.1.2 Introduction to the Process of This Method

In recent years, the drug resistance state of tumors, regarded as an important process in the evolution of tumor resistance, has been well studied by using a variety of machine learning methods. This type of methods analyze the drug resistance mechanism activated by tumor cell lines in different drug resistance state. They mainly consist of four steps. 1) The first step is to analyze the drug resistance mechanism produced from the tumor drug resistance state. This step requires the construction of a time-series biological network (T-BioPPI). T-BioPPI integrates multiple database biological association networks to form a relatively comprehensive biological protein interaction network of BioPPI, in which the LINCS database provides tumor cell line gene expression profile data at three time points. The data at these time points are analyzed for differentially expressed genes. Each gene is expressed as a different node between time layers in the biological network. At this point, the gene interaction networks at various time points are linked together to construct a large time-series biological network, thus identifying the important genes in the biological network through regression analysis. Because each drug has a relatively fixed target in the cells, the key lies in the action time of the drug. Using the drug dose as an independent variable and gene expression as a dependent variable, regression analysis of gene expression over time in tumor cells following drug treatment can be performed. Following this step is an initiated step to map the genes of each tumor cell line to

T-BioPPI. Then, by restarting the random walk, an analysis of walking from the nodes in the 0-h network to the 24-h network can be performed. This is followed by dimensionality reduction and k-means++ clustering. The evolution prediction model is the classification result of tumor cell lines obtained after clustering. 2) Second, a deep learning model is established for predicting the evolution of the tumor drug resistance state. In general, long short-term memory networks are applied to assess the evolution of tumor drug resistance through ordered sequences of gene expression changes. During this step, the classified cell line data are used as the basis for long short-term memory model construction, and the gene expression profile data of similar tumor cell lines are used as the training set for deep learning data. 3) The third step is to design drug application strategies that interfere with tumor drug resistance to cover single-target drug identification and combination drug identification. 4) The fourth step is to verify the predicted drug resistance evolution model and medication plan through cytology and gene chip experiments. Three main sub-steps are involved. The first sub-step is the cultivation of drug-resistant cell lines, the second sub-step involves gene chip experiments, and the last sub-step is the cell proliferation inhibition test. **Figure 2** presents a reference for the drug resistance evolution model based on the tumor drug resistance state and the general process of the drug administration strategy.

### 2.2 Graph Embedding

Computational theory tools such as graph representation learning are widely used to establish standard data sets and online databases for anti-cancer drug sensitivity mutation data using. Then analysis models are established to conduct in-depth research and exploration on anti-cancer drug sensitivity mutation prediction methods. The project mainly studied the following aspects. 1) First, the construction of anti-cancer drug sensitivity mutation databases. After collecting anti-cancer drug sensitivity mutation data from multiple cancer genome projects, a document classifier for cancer-related mutations can be constructed by using machine learning text mining technology. The classifier intends to facilitate access to the literature about cancer–mutation–drug information in the PubMed database. After obtaining these related documents, professional personnel can collect and annotate the relevant anti-cancer drug sensitivity mutation information. When mutation data information is obtained for the first time, standard tools are employed to organize the information annotations of each entry into a standard format. Then, the obtained original data sets are integrated with the source of literature mining. Finally, a user-friendly anti-cancer drug sensitivity mutation database web interface based on the Browser/Server model is developed, which is open for users to view and download data. 2) The second aspect focuses on research on the prediction method of drug sensitivity markers. According to the characteristics of drug sensitivity-related mutations, known anti-cancer drug sensitivity mutation data is sorted and preprocessed. Then, existing feature quantification methods are collected and well analyzed. Meanwhile, wild-type and mutant DNA sequences are used for feature quantification to permit the information before and after the appearance of the mutation, which can further ensure the reliability of the results. The background network is obtained by calculating the mutation–mutation similarity. 3) Third, a drug–drug network is



extracted from the research on the prediction method of mutation–drug interaction pairs. Then, a multi-source heterogeneous drug interaction network is established through techniques such as similar network fusion. 4) Finally, the graph representation learning method is adopted to predict the relationship of the mutation–drug interaction pair and develop corresponding prediction software and an online prediction platform. The flowchart of graph embedding-based algorithm NEDTP is shown in **Figure 3**.

### 2.3 Capsule Network and Shapley Value Method

Deep learning has shown impressive performance in many tasks (Jiang et al., 2013; Guo et al., 2020; Jin et al., 2020; Tao et al., 2020; Yu et al., 2020b; Zhang et al., 2020b; Zhao et al., 2020; Jin et al., 2021; Liu et al., 2021b; Lv et al., 2021; Su et al., 2021; Wang et al., 2021a; Xu et al., 2021; Yu et al., 2021). This deep model-based strategy intends to build a deep feed-forward network and drug fingerprint encoding method to obtain the disease cell lines and drug quantitative characteristics. To identify the direct correlation between drug groups and disease-specific gene expression profiles, this project adopted the capsule network and the encoder–decoder model of the attention mechanism to predict the sensitivity of cancer cell lines to single and combination drug regimens. Capsule network is an improved convolutional neural network, which loads the information of feature states learned in the network into capsules in the form of vectors. The capsule preserves precise information about position and posture, making the visual entity locally invariant. While traditional deep learning methods output as a single scalar on a single neuron, and realize the invariance of perspective through maximum pooling method, it loses a lot of valuable information and fails to take into account the relative spatial relationship between coding features. The capsule network can learn the posture information of different cells from the cancer cell line, and convert the information that might be missed by the traditional CNN network into high-level features, which can be used to predict the sensitivity of drugs to the cancer cell line. The contribution/inhibition relationship between drug groups for specific diseases was obtained using the Shapley value method of cooperative game theory to analyze the convolutional neural network model (Aumann and Shapley, 1971; Karim et al., 2019; Cai et al., 2020a; Cai et al., 2020b; Mo et al., 2020), as well as through calculation and comparative analysis of the marginal contributions between drug



groups. The project mainly studied three points: 1) Feature quantification methods oriented at the prior knowledge of genomics and drug targeting information to construct a deep feed-forward network and encode drug fingerprints based on drug targeting relationships. 2) Designing a deep learning model adapted to gene expression profile and drug gene data. First a network structure is built to analyze the basic structural association relationship between drug groups and gene expression profiles. And the capsule network is used to extract the characteristics of the cancer cell line and the drug itself. Moreover, the encoder–decoder model of the attention mechanism is adopted for the fusion of heterogeneous features. 3) Constructing a gene expression profile–drug group network by cooperative game model. Then the gene–drug group network is applied to calculate the enhancement/inhibition degree of the drug fingerprint and identify the set with obvious enhancement/inhibition effects in the drug group. This type of method considers both the enhancement/inhibition relationship between drug combinations. It first integrates the drug combination data from various sources, then extracts the enhancement/inhibition relationship combinations of different drug combinations, and finally predicts the sensitivity of drug combination with machine learning algorithms based on the different feature combinations.

## 2.4 Drug Resistance Association Analysis

The strategy mainly aims to establish a recombinant strain drug resistance analysis platform, and verify the hypothesis related to recombinant drug resistance by targeting circulating recombinant forms (CRFs) (Ru et al., 2020; Wang et al., 2020a). Data related to drug resistance can be obtained through appropriate and efficient data mining methods. By combining SeqFeatR and Bayesian factor methods, complex hierarchical models can be used to quantify drug combinations (Plummer, 2003; Bettina et al., 2016; Zhao et al., 2019; Hu et al., 2020; Zeng et al., 2020a; Zeng et al., 2020b; Hu et al., 2021a; Hu et al., 2021b; Song et al., 2021). Meanwhile, reliable associations can be identified for recombinant HIV-1 for application in anti-viral therapy based on the link between base substitutions in viral sequences and the viral genomic background. The direct coupling analysis method is used to predict the interaction between the associated mutations in the protein and analyze the nearest neighbor between the sites associated with drug resistance. The main research contents are as follows: 1) establish a drug resistance association analysis platform for HIV-1 CRFs and 2) propose and verify that the HIV-1 CRFs are related to drug resistance mutations. This method mainly aims to establish a recombinant HIV-1 drug resistance analysis platform and spread it to other recombinant pandemic areas across the globe. It initially involves data and model inference. In this project, HIV-1 pol serves as the research object, and model inference is achieved through JRip of the RWeka software package of the R system, which makes fast rule inference on the aforementioned three sets of data by adopting the RIPPER algorithm. Then, reliability verification of the model inference (leave-one-out classification verification) is conducted to obtain the statistical evaluation of the rule inference results (Zeng et al., 2018; Zeng et al., 2019; Dao et al., 2020; Fu et al., 2020; Zulfiqar et al., 2021). Finally, the sequence characteristics of recombinant HIV-1, recombinant drug-related patterns, and recombinant characteristic

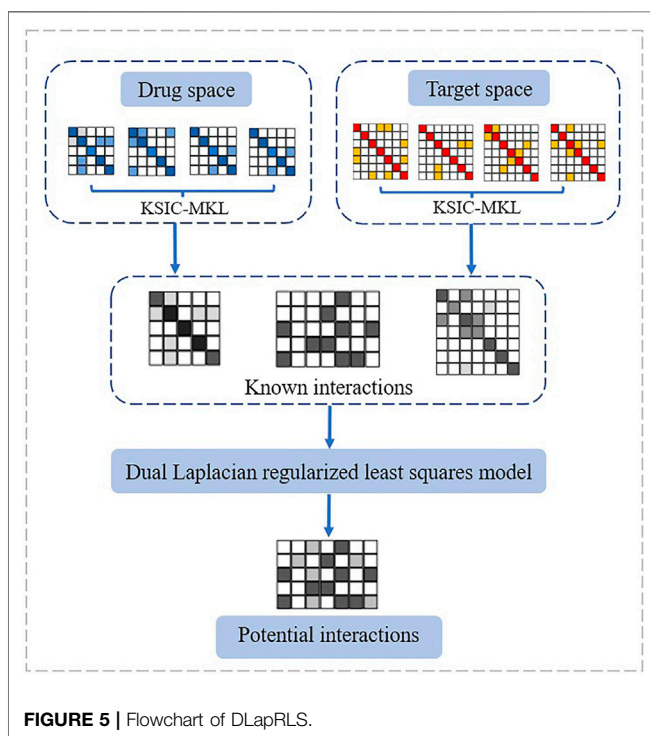
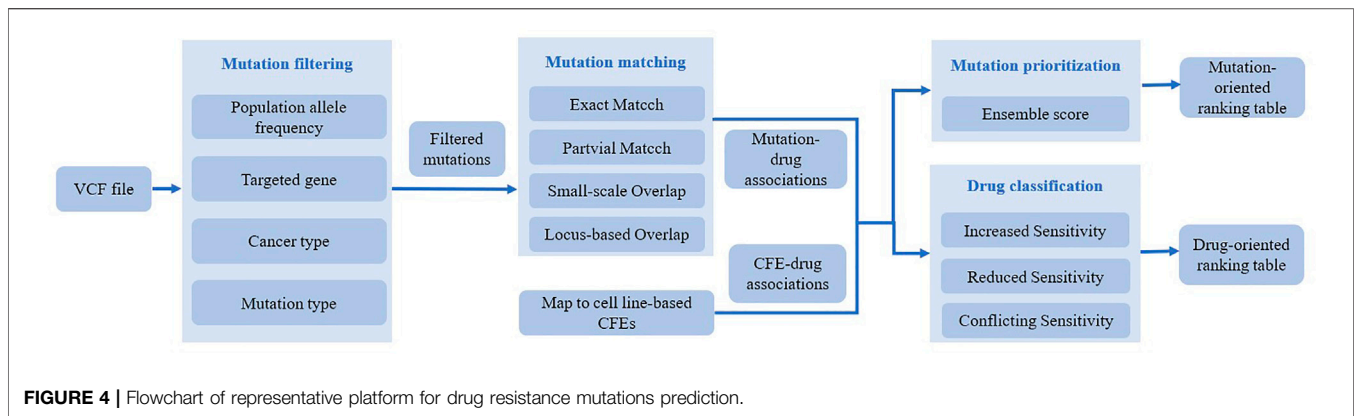
drug resistance patterns will be obtained. In general, this method performs drug resistance information interpretation and correlation analysis of HIV-1 recombinant characteristic drug resistance mutations, and then performs computer modeling and construction verification. The flowchart of representative platform for drug resistance mutations prediction is shown in **Figure 4**.

## 2.5 Summary

This chapter mainly discusses the significance and indispensability of drug sensitivity in bioinformatics research by introducing different methods. A persistent problem has arisen in the research and development of anti-cancer drugs and in the medical and health fields, namely the issue of tumor resistance. Methods reviewed in this study can assist with the prediction of drug sensitivity. They encompass oncology, pharmacy, and computer science, and strategies to predict tumor resistance, design rational drug strategies, and construct computer models were principally covered. These methods have promoted the research and development of bioinformatic fields such as computational methodology and algorithm design. Moreover, research reviewed in this study can be directly applied to anti-cancer precision medicine, new drug identification, and other systems. They have exhibited broad market application prospects and a further possibility to improve the effectiveness of AIDS treatment and lower the cost of its prevention and control.

## 3 LITERATURE CONTRIBUTION

Tumor drug tolerance is an important process in the evolution of tumor drug resistance, and the drug resistance mechanism activated by the tumor in the drug resistance state remains unclear. Bioinformatic analysis and research on mutations associated with anti-cancer drug sensitivity are expanding. Hopefully, the methods reviewed in this study will contribute to overcoming existing problems. They analyzed the drug resistance mechanism of tumor cell lines in a drug-resistant state based on a variety of machine learning (Wei et al., 2014; Wei et al., 2017a; Wei et al., 2017b; Ding et al., 2020a; Ding et al., 2020b; Wang et al., 2020b; Wang et al., 2021b) and deep learning methods to study the gene expression profiles of a large number of drug-resistant tumor cells (Lv et al., 2019; Su et al., 2019; He et al., 2020; Li et al., 2020; Peng et al., 2020; Su et al., 2020; Zhang et al., 2020c; Cui et al., 2021). The flowchart of a representative method DLapRLS is shown in **Figure 5**. The established prediction model provides a new strategy for future research on tumor drug resistance. Cell experiments are also applied to block the evolution of drug resistance in the tumor drug resistance state by using single drugs and combination regimens. A new random walk-based graph representation learning algorithm was proposed to the predict of anti-cancer drug sensitivity mutation data. It incorporates gene–drug interaction network information into the node representation of mutation–mutation networks for the comprehensive and systematic command of the inherent properties of such mutations. Moreover, a mutation–drug network graph representation algorithm with multi-source



heterogeneous information was developed to predict mutations associated with anti-cancer drug sensitivity and sensitivity/resistance to specific drugs. Meanwhile, a network prediction platform available for researchers was also developed (Zhang et al., 2020a). An exploratory method based on the Shapley value of cooperative game theory was proposed to analyze the convolutional neural network model. Through the differential analysis of the contributions of monotherapies and synergistic drug combinations in specific disease cell lines, candidate drug group collection of the enhancement/inhibition relation in drug components was obtained. Further application and promotion of the drug resistance association analysis platform can provide strategies for controlling the spread of circulating recombinant drug-resistant HIV-1 strain. At the same time, it can also help reduce the blind use of drugs against circulating recombinant drug-

resistant HIV-1 strain, improve the treatment effectiveness, and lower the cost of prevention and control of AIDS.

## 4 CONCLUSION

In recent years, anti-cancer drugs have consistently been the focus of new drug development. Methods that can accurately predict drug sensitivity are urgently needed to facilitate drug development and disease prevention in the field of biomedical health. In this study, a comprehensive review was provided concerning the analysis of the drug resistance mechanisms of tumor cell lines in the drug-resistant state by using a variety of machine learning methods. The two-step cancer–mutation–drug triad prediction is achieved through text mining technology based on machine learning. It has laid solid foundations for the subsequent construction and update of the drug sensitivity mutation database through a combination of manual annotations. The wide distribution of HIV-1 recombinant types and the formation of drug-resistant strains will facilitate our study of recombinant characteristic drug resistance patterns and drug resistance associations. The aforementioned summary indicates the necessity of the current review.

## AUTHOR CONTRIBUTIONS

YC, XL and LS conceived this work. YC and LJ collected papers and data, studied the literature, and performed the analysis. YC, LJ, XL and LS wrote, revised, and approved the manuscript.

## FUNDING

The work was supported by the Special Science Foundation of Quzhou (2021D004).

## ACKNOWLEDGMENTS

We thank Joe Barber Jr., from Liwen Bianji (Edanz) ([www.liwenbianji.cn/](http://www.liwenbianji.cn/)), for editing the English text of a draft of this manuscript.

## REFERENCES

- An, Q., and Yu, L. (2021). A Heterogeneous Network Embedding Framework for Predicting Similarity-Based Drug-Target Interactions. *Brief. Bioinformatics* 22, 1. doi:10.1093/bib/bbab275
- Aumann, R. J., and Shapley, L. S. (1971). *Values of Non-atomic Games, Part V. RAND Corporation*. Santa Monica, California: RAND.
- Azad, M. T. A., Qulsum, U., and Tsukahara, T. (2019). Comparative Activity of Adenosine Deaminase Acting on RNA (ADARs) Isoforms for Correction of Genetic Code in Gene Therapy. *Curr. Gene Ther.* 19 (1), 31–39. doi:10.2174/1566523218666181114122116
- Bettina, B., Jorg, T., and Daniel, H. (2016). SeqFeatR for the Discovery of Feature-Sequence Associations. *PLOS ONE* 11 (1), e0146409. doi:10.1371/journal.pone.0146409
- Cai, L., Ren, X., Fu, X., Peng, L., Gao, M., Zeng, X., et al. (2020). iEnhancer-XG: Interpretable Sequence-Based Enhancers and Their Strength Predictor. *Bioinformatics* 37 (8), 1060–1067. doi:10.1093/bioinformatics/btaa914
- Cai, L., Wang, L., Fu, X., Xia, C., Zeng, X., Zou, Q., et al. (2020). ITP-pred: an Interpretable Method for Predicting Therapeutic Peptides with Fused Features Low-Dimension Representation. *Brief. Bioinform.* 22 (4), bbab367. doi:10.1093/bib/bbaa367
- Carr, T. H., McEwen, R., Dougherty, B., Johnson, J. H., Dry, J. R., Lai, Z., et al. (2016). Defining Actionable Mutations for Oncology Therapeutic Development. *Nat. Rev. Cancer* 16 (5), 319–329. doi:10.1038/nrc.2016.35
- Castro-Nallar, E., Pérez-Losada, M., Burton, G. F., and Crandall, K. A. (2012). The Evolution of HIV: Inferences Using Phylogenetics. *Mol. Phylogenet. Evol.* 62 (2), 777–792. doi:10.1016/j.ympev.2011.11.019
- Cheng, L., Hu, Y., Sun, J., Zhou, M., and Jiang, Q. (2018). DincRNA: a Comprehensive Web-Based Bioinformatics Toolkit for Exploring Disease Associations and ncRNA Function. *Bioinformatics* 34 (11), 1953–1956. doi:10.1093/bioinformatics/bty002
- Cheng, L., Qi, C., Yang, H., Lu, M., Cai, Y., Fu, T., et al. (2021). gutMGene: a Comprehensive Database for Target Genes of Gut Microbes and Microbial Metabolites. *Nucleic Acids Res.* 1, gkab786. doi:10.1093/nar/gkab786
- Cheng, L., Qi, C., Zhuang, H., Fu, T., and Zhang, X. (2020). gutMDisorder: a Comprehensive Database for Dysbiosis of the Gut Microbiota in Disorders and Interventions. *Nucleic Acids Res.* 48 (D1), D554–D560. doi:10.1093/nar/gkz843
- Cheng, L., Yang, H., Zhao, H., Pei, X., Shi, H., Sun, J., et al. (2019). MetSigDis: a Manually Curated Resource for the Metabolic Signatures of Diseases. *Brief Bioinform* 20 (1), 203–209. doi:10.1093/bib/bbx103
- Cui, F., Zhang, Z., and Zou, Q. (2021). Sequence Representation Approaches for Sequence-Based Protein Prediction Tasks that Use Deep Learning. *Brief. Funct. Genomics* 20 (1), 61–73. doi:10.1093/bfpgp/ela030
- Dao, F. Y., Lv, H., Zulfiqar, H., Yang, H., Su, W., Gao, H., et al. (2020). A Computational Platform to Identify Origins of Replication Sites in Eukaryotes. *Brief Bioinform* 22 (2), 1940–1950. doi:10.1093/bib/bbaa017
- Ding, Y., Tang, J., and Guo, F. (2020). Identification of Drug-Target Interactions via Dual Laplacian Regularized Least Squares with Multiple Kernel Fusion. *Knowledge-Based Syst.* 204, 106254. doi:10.1016/j.knsys.2020.106254
- Ding, Y., Tang, J., and Guo, F. (2020). Identification of Drug-Target Interactions via Fuzzy Bipartite Local Model. *Neural Comput. Applic* 32, 10303–10319. doi:10.1007/s00521-019-04569-z
- Fu, X., Cai, L., Zeng, X., and Zou, Q. (2020). StackCPPred: a Stacking and Pairwise Energy Content-Based Prediction of Cell-Penetrating Peptides and Their Uptake Efficiency. *Bioinformatics* 36 (10), 3028–3034. doi:10.1093/bioinformatics/btaa131
- Guo, Z., Wang, P., Liu, Z., and Zhao, Y. (2020). Discrimination of Thermophilic Proteins and Non-thermophilic Proteins Using Feature Dimension Reduction. *Front. Bioeng. Biotechnol.* 8, 584807. doi:10.3389/fbioe.2020.584807
- Hangauer, M. J., Viswanathan, V. S., Ryan, M. J., Bole, D., Eaton, J. K., Matov, A., et al. (2017). Drug-tolerant Persister Cancer Cells Are Vulnerable to GPX4 Inhibition. *Nature* 551 (7679), 247–250. doi:10.1038/nature24297
- Hanna, N. H. (2006). EGFR Mutation and Resistance of Non-small-cell Lung Cancer to Gefitinib. *Yearb. Oncol.* 2006, 228–229. doi:10.1016/s1040-1741(08)70161-1
- He, B., Zhu, R., Yang, H., Lu, Q., Wang, W., Song, L., et al. (2020). Assessing the Impact of Data Preprocessing on Analyzing Next Generation Sequencing Data. *Front. Bioeng. Biotechnol.* 8, 817. doi:10.3389/fbioe.2020.00817
- Hemelaar, J. (2012). The Origin and Diversity of the HIV-1 Pandemic. *Trends Mol. Med.* 18 (3), 182–192. doi:10.1016/j.molmed.2011.12.001
- Hu, Y., Qiu, S., and Cheng, L. (2021). Integration of Multiple-Omics Data to Analyze the Population-specific Differences for Coronary Artery Disease. *Comput. Math. Methods Med.* 2021, 7036592. doi:10.1155/2021/7036592
- Hu, Y., Sun, J.-Y., Zhang, Y., and Zhang, H. (2021). rs1990622 Variant Associates with Alzheimer's Disease and Regulates TMEM106B Expression in Human Brain Tissues. *BMC Med.* 19 (1), 11. doi:10.1186/s12916-020-01883-5
- Hu, Y., Zhang, H., Liu, B., Gao, S., Wang, T., Han, Z., et al. (2020). rs34331204 Regulates TSPAN13 Expression and Contributes to Alzheimer's Disease with Sex Differences. *Brain* 143 (11), e95. doi:10.1093/brain/awaa302
- Jaiswal, A., AbdAlmageed, W., Wu, Y., and Natarajan, P. (2018). CapsuleGAN: Generative Adversarial Capsule Network. Available at: <https://arxiv.org/abs/1802.06167>.
- Jennifer, H., Kim, J., Yadav, V., Amato, C., Robinson, S. E., Seelenfreund, E., et al. (2016). IMPACT: a Whole-Exome Sequencing Analysis Pipeline for Integrating Molecular Profiles with Actionable Therapeutics in Clinical Samples. *J. Am. Med. Inform. Assoc.* 23 (4), 721. doi:10.1093/jamia/ocw022
- Jiang, Q., Wang, G., Jin, S., Li, Y., and Wang, Y. (2013). Predicting Human microRNA-Disease Associations Based on Support Vector Machine. *Int. J. Data Min Bioinform* 8 (3), 282–293. doi:10.1504/ijdm.2013.056078
- Jin, Q., Cui, H., Sun, C., Meng, Z., and Su, R. (2021). Free-form Tumor Synthesis in Computed Tomography Images via Richer Generative Adversarial Network. *Knowledge-Based Syst.* 218, 106753. doi:10.1016/j.knsys.2021.106753
- Jin, S., Zeng, X., Xia, F., Huang, W., and Liu, X. (2020). Application of Deep Learning Methods in Biological Networks. *Brief Bioinform* 22 (2), 1902–1917. doi:10.1093/bib/bbaa043
- Karim, M. R., Cochez, M., Beyan, O., and Decker, S., OncoNetExplainer: Explainable Predictions of Cancer Types Based on Gene Expression Data. in Proceedings of the 2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE). Athens, Greece October 2019 2019. doi:10.1109/bibe.2019.00081
- Li, J., Pu, Y., Tang, J., Zou, Q., and Guo, F. (2020). DeepAVP: a Dual-Channel Deep Neural Network for Identifying Variable-Length Antiviral Peptides. *IEEE J. Biomed. Health Inform.* 24 (10), 3012–3019. doi:10.1109/JBHI.2020.2977091
- Li, M. J., Yao, H., Huang, D., Liu, H., Liu, Z., Xu, H., et al. mTCTScan: a Comprehensive Platform for Annotation and Prioritization of Mutations Affecting Drug Sensitivity in Cancers. *Nucleic Acids Res. Oxford, United Kingdom* 2017 45p. W215–W221. doi:10.1093/nar/gkx400
- Liu, C., Wei, D., Xiang, J., Ren, F., Huang, L., Lang, J., et al. (2020). An Improved Anticancer Drug-Response Prediction Based on an Ensemble Method Integrating Matrix Completion and Ridge Regression. *Mol. Ther. Nucleic Acids* 21, 676–686. doi:10.1016/j.omtn.2020.07.003
- Liu, H., Qiu, C., Wang, B., Bing, P., Tian, G., Zhang, X., et al. (2021). Evaluating DNA Methylation, Gene Expression, Somatic Mutation, and Their Combinations in Inferring Tumor Tissue-Of-Origin. *Front. Cel Dev Biol* 9, 619330. doi:10.3389/fcell.2021.619330
- Liu, J., Su, R., Zhang, J., and Wei, L. (2021). Classification and Gene Selection of Triple-Negative Breast Cancer Subtype Embedding Gene Connectivity Matrix in Deep Neural Network. *Brief. Bioinform.* 22 (5), bbab395. doi:10.1093/bib/bbaa395
- Liu, K., and Chen, W. (2020). iMRM: a Platform for Simultaneously Identifying Multiple Kinds of RNA Modifications. *Bioinformatics* 36 (11), 3336–3342. doi:10.1093/bioinformatics/btaa155
- Liu, X., Yang, J., Zhang, Y., Fang, Y., Wang, F., Wang, J., et al. (2016). A Systematic Study on Drug-Response Associated Genes Using Baseline Gene Expressions of the Cancer Cell Line Encyclopedia. *Sci. Rep.* 6, 22811. doi:10.1038/srep22811
- Lv, H., Dao, F.-Y., Zulfiqar, H., and Lin, H. (2021). DeepIPs: Comprehensive Assessment and Computational Identification of Phosphorylation Sites of SARS-CoV-2 Infection Using a Deep Learning-Based Approach. *Brief Bioinform* 22 (6), bbab244. doi:10.1093/bib/bbab244
- Lv, Z., Ao, C., and Zou, Q. (2019). Protein Function Prediction: From Traditional Classifier to Deep Learning. *Proteomics* 19 (14), e1900119. doi:10.1002/pmic.201900119
- Mo, F., Luo, Y., Fan, D., Zeng, H., Zhao, Y., Luo, M., et al. (2020). Integrated Analysis of mRNA-Seq and miRNA-Seq to Identify C-MYC, YAP1 and miR-3960 as Major Players in the Anticancer Effects of Caffeic Acid Phenethyl Ester in Human Small Cell Lung Cancer Cell Line. *Curr. Gene Ther.* 20 (1), 15–24. doi:10.2174/1566523220666200523165159
- O'Donnell, J., Teng, M., and Smyth, M. J. (2018). Cancer Immunoediting and Resistance to T Cell-Based Immunotherapy. *Nat. Rev. Clin. Oncol.* 16 (3), 151–167.
- Oxnard, G. R., and Geoffrey, R. (2016). The Cellular Origins of Drug Resistance in Cancer. *Nat. Med.* 22 (3), 232–234. doi:10.1038/nm.4058



- Peng, L., Zhou, D., Liu, W., Zhou, L., Wang, L., Zhao, B., et al. (2020). Prioritizing Human Microbe-Disease Associations Utilizing a Node-Information-Based Link Propagation Method. *IEEE Access* 8, 31341–31349. doi:10.1109/access.2020.2972283
- Plummer, M., JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling, Proceedings of the 3rd international workshop on Distributed Statistical Computing Vienna Austria March 2003 2003.
- Qi, C., Wang, C., Zhao, L., Zijun, Z., Wang, P., Zhang, S., et al. (2021). SCovid: Single-Cell Atlases for Exposing Molecular Characteristics of COVID-19 across 10 Human Tissues. *Nucleic Acids Res.* 1, gkab881, 2021. Oxford, United Kingdom. doi:10.1093/nar/gkab881
- Recasens, A., and Munoz, L. (2019). Targeting Cancer Cell Dormancy. *Trends Pharmacol. Sci.* 40 (2), 128–141. doi:10.1016/j.tips.2018.12.004
- Restifo, N. P., Smyth, M. J., and Snyder, A. (2016). Acquired Resistance to Immunotherapy and Future Challenges. *Nat. Rev. Cancer* 16 (2), 121–126. doi:10.1038/nrc.2016.2
- Ru, X., Wang, L., Li, L., Ding, H., Ye, X., and Zou, Q. (2020). Exploration of the Correlation between GPCRs and Drugs Based on a Learning to Rank Algorithm. *Comput. Biol. Med.* 119, 103660. doi:10.1016/j.compbiomed.2020.103660
- Russo, A. (2015). *Targeted Therapies for Non-small Cell Lung Cancer*. New York: Springer.
- Schmitt, M. W., Loeb, L. A., and Salk, J. J. (2016). The Influence of Subclonal Resistance Mutations on Targeted Cancer Therapy. *Nat. Rev. Clin. Oncol.* 13, 335. doi:10.1038/nrclinonc.2015.175
- Shang, Y., Gao, L., Zou, Q., and Yu, L. (2021). Prediction of Drug-Target Interactions Based on Multi-Layer Network Representation Learning. *Neurocomputing* 434, 80–89. doi:10.1016/j.neucom.2020.12.068
- Shaw, G. M., and Hunter, E. (2012). HIV Transmission. *Cold Spring Harb Perspect. Med.* 2 (11). doi:10.1101/cshperspect.a006965
- Song, B., Li, F., Liu, Y., and Zeng, X. (2021). Deep Learning Methods for Biomedical Named Entity Recognition: a Survey and Qualitative Comparison. *Brief. Bioinform.* 22, bbab282. doi:10.1093/bib/bbab282
- Song, B., Li, K., Orellana-Martín, D., Valencia-Cabrera, L., and Pérez-Jiménez, M. J. (2020). Cell-like P Systems with Evolutionary Symport/Antiport Rules and Membrane Creation. *Inf. Comput.* 275, 104542. doi:10.1016/j.ic.2020.104542
- Su, R., Hu, J., Zou, Q., Manavalan, B., and Wei, L. (2020). Empirical Comparison and Analysis of Web-Based Cell-Penetrating Peptide Prediction Tools. *Brief Bioinform.* 21 (2), 408–420. doi:10.1093/bib/bby124
- Su, R., Liu, X., Jin, Q., Liu, X., and Wei, L. (2021). Identification of Glioblastoma Molecular Subtype and Prognosis Based on Deep MRI Features. *Knowledge-Based Syst.* 232, 107490. doi:10.1016/j.knsys.2021.107490
- Su, R., Liu, X., Wei, L., and Zou, Q. (2019). Deep-Resp-Forest: A Deep forest Model to Predict Anti-cancer Drug Response. *Methods* 166, 91–102. doi:10.1016/j.ymeth.2019.02.009
- Tao, Z., Li, Y., Teng, Z., and Zhao, Y. (2020). A Method for Identifying Vesicle Transport Proteins Based on LibSVM and MRMD. *Comput. Math. Methods Med.* 2020, 8926750. doi:10.1155/2020/8926750
- Wang, D., Zhang, Z., Jiang, Y., Mao, Z., Wang, D., Lin, H., et al. (2021). DM3Loc: Multi-Label mRNA Subcellular Localization Prediction and Analysis Based on Multi-Head Self-Attention Mechanism. *Nucleic Acids Res.* 49 (8), e46. doi:10.1093/nar/gkab016
- Wang, H., Ding, Y., Tang, J., and Guo, F. (2020). Identification of Membrane Protein Types via Multivariate Information Fusion with Hilbert-Schmidt Independence Criterion. *Neurocomputing* 383, 257–269. doi:10.1016/j.neucom.2019.11.103
- Wang, H., Tang, J., Ding, Y., and Guo, F. (2021). Exploring Associations of Non-coding RNAs in Human Diseases via Three-Matrix Factorization with Hypergraph-Regular Terms on center Kernel Alignment. *Brief. Bioinform.* 22, 1. doi:10.1093/bib/bbaa409
- Wang, J., Shi, Y., Wang, X., and Chang, H. (2020). A Drug Target Interaction Prediction Based on LINE-RF Learning. *Cbio* 15 (7), 750–757. doi:10.2174/1574893615666191227092453
- Wei, L., Liao, M., Gao, Y., Ji, R., He, Z., and Zou, Q. (2014). Improved and Promising Identification of Human MicroRNAs by Incorporating a High-Quality Negative Set. *Ieee/acm Trans. Comput. Biol. Bioinform.* 11 (1), 192–201. doi:10.1109/TCBB.2013.146
- Wei, L., Wan, S., Guo, J., and Wong, K. K. (2017). A Novel Hierarchical Selective Ensemble Classifier with Bioinformatics Application. *Artif. Intell. Med.* 83, 82–90. doi:10.1016/j.artmed.2017.02.005
- Wei, L., Xing, P., Zeng, J., Chen, J., Su, R., and Guo, F. (2017). Improved Prediction of Protein-Protein Interactions Using Novel Negative Samples, Features, and an Ensemble Classifier. *Artif. Intell. Med.* 83, 67–74. doi:10.1016/j.artmed.2017.03.001
- Xu, Z., Luo, M., Lin, W., Xue, G., Wang, P., Jin, X., et al. (2021). DLpTCR: an Ensemble Deep Learning Framework for Predicting Immunogenic Peptide Recognized by T Cell Receptor. *Brief Bioinform.* 22 (6). doi:10.1093/bib/bbab335
- Yu, L., Xu, F., and Gao, L. (2020). Predict New Therapeutic Drugs for Hepatocellular Carcinoma Based on Gene Mutation and Expression. *Front. Bioeng. Biotechnol.* 8, 8. doi:10.3389/fbioe.2020.00008
- Yu, L., Xia, M., and An, Q. (2021). A Network Embedding Framework Based on Integrating Multiplex Network for Drug Combination Prediction. *Brief. Bioinformatics* 364. doi:10.1093/bib/bbab364
- Yu, L., Zhou, D., Gao, L., and Zha, Y. (2020). Prediction of Drug Response in Multilayer Networks Based on Fusion of Multiomics Data. *Methods* 192, 85–92. doi:10.1016/j.ymeth.2020.08.006
- Zeng, X., Liu, L., Lü, L., and Zou, Q. (2018). Prediction of Potential Disease-Associated microRNAs Using Structural Perturbation Method. *Bioinformatics* 34 (14), 2425–2432. doi:10.1093/bioinformatics/bty112
- Zeng, X., Song, X., Ma, T., Pan, X., Zhou, Y., Hou, Y., et al. (2020). Repurpose Open Data to Discover Therapeutics for COVID-19 Using Deep Learning. *J. Proteome Res.* 19 (11), 4624–4636. doi:10.1021/acs.jproteome.0c00316
- Zeng, X., Zhu, S., Liu, X., Zhou, Y., Nussinov, R., and Cheng, F. (2019). deepDR: a Network-Based Deep Learning Approach to In Silico Drug Repositioning. *Bioinformatics* 35 (24), 5191–5198. doi:10.1093/bioinformatics/btz418
- Zeng, X., Zhu, S., Lu, W., Liu, Z., Huang, J., Zhou, Y., et al. (2020). Target Identification Among Known Drugs by Deep Learning from Heterogeneous Networks. *Chem. Sci.* 11 (7), 1775–1797. doi:10.1039/c9sc04336e
- Zhang, L., Xiao, X., and Xu, Z. C. (2020). iPromoter-5mC: A Novel Fusion Decision Predictor for the Identification of 5-Methylcytosine Sites in Genome-wide DNA Promoters. *Front. Cell Dev Biol.* 8, 614. doi:10.3389/fcell.2020.00614
- Zhang, S., Su, M., Sun, Z., Lu, H., and Zhang, Y. (2020). The Signature of Pharmaceutical Sensitivity Based on ctDNA Mutation in Eleven Cancers. *Exp. Biol. Med. (Maywood)* 245 (8), 720–732. doi:10.1177/1535370220906518
- Zhang, Y., Yang, J., Chen, S., and Gong, M. (2020). Review of the Applications of Deep Learning in Bioinformatics. *Curr. Bioinformatics* 15 (8), 898–911. doi:10.2174/1574893615999200711165743
- Zhao, T., Wang, D., Hu, Y., Zhang, N., Zang, T., and Wang, Y. (2019). Identifying Alzheimer's Disease-Related miRNA Based on Semi-clustering. *Curr. Gene Ther.* 19 (4), 216–223. doi:10.2174/1566523219666190924113737
- Zhao, W., Ye, J., Yang, M., and Lei, Z., Investigating Capsule Networks with Dynamic Routing for Text Classification. in Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium November 2018 2018. doi:10.18653/v1/D18-1350
- Zhao, X., Jiao, Q., Li, H., Wu, Y., Wang, H., Huang, S., et al. (2020). ECFS-DEA: an Ensemble Classifier-Based Feature Selection for Differential Expression Analysis on Expression Profiles. *BMC Bioinformatics* 21 (1), 43. doi:10.1186/s12859-020-3388-y
- Zhuang, J., Dai, S., Zhang, L., and Gao, P. (2020). Identifying Breast Cancer-Induced Gene Perturbations and its Application in Guiding Drug Repurposing. *Curr. Bioinformatics* 15 (9), 1075–1089.
- Zulficar, H., Yuan, S.-S., Huang, Q.-L., Sun, Z.-J., Dao, F.-Y., Yu, X.-L., et al. (2021). Identification of Cyclin Protein Using Gradient Boost Decision Tree Algorithm. *Comput. Struct. Biotechnol. J.* 19, 4123–4131. doi:10.1016/j.csbj.2021.07.013

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Chen, Juan, Lv and Shi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.