



A Review of Current *In Silico* Methods for Repositioning Drugs and Chemical Compounds

Binsheng He^{1*†}, Fangxing Hou^{2†}, Changjing Ren^{3,4†}, Pingping Bing^{1*} and Xiangzuo Xiao^{5*}

¹ Academician Workstation, Changsha Medical University, Changsha, China, ² Queen Mary School, Nanchang University, Jiangxi, China, ³ School of Science, Dalian Maritime University, Dalian, China, ⁴ Genies Beijing Co., Ltd., Beijing, China, ⁵ Department of Radiology, The First Affiliated Hospital of Nanchang University, Jiangxi, China

OPEN ACCESS

Edited by:

Jing Sun,
George Washington University,
United States

Reviewed by:

Taigang Liu,
Shanghai Ocean University, China
Peizhen Wang,
Anhui University of Technology, China

*Correspondence:

Xiangzuo Xiao
xxz8785@sina.com
Binsheng He
hbcsmu@163.com
Pingping Bing
bpping@163.com

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Cancer Genetics,
a section of the journal
Frontiers in Oncology

Received: 18 May 2021

Accepted: 07 July 2021

Published: 22 July 2021

Citation:

He B, Hou F, Ren C, Bing P and Xiao X
(2021) A Review of Current *In Silico*
Methods for Repositioning Drugs
and Chemical Compounds.
Front. Oncol. 11:711225.
doi: 10.3389/fonc.2021.711225

Drug repositioning is a new way of applying the existing therapeutics to new disease indications. Due to the exorbitant cost and high failure rate in developing new drugs, the continued use of existing drugs for treatment, especially anti-tumor drugs, has become a widespread practice. With the assistance of high-throughput sequencing techniques, many efficient methods have been proposed and applied in drug repositioning and individualized tumor treatment. Current computational methods for repositioning drugs and chemical compounds can be divided into four categories: (i) feature-based methods, (ii) matrix decomposition-based methods, (iii) network-based methods, and (iv) reverse transcriptome-based methods. In this article, we comprehensively review the widely used methods in the above four categories. Finally, we summarize the advantages and disadvantages of these methods and indicate future directions for more sensitive computational drug repositioning methods and individualized tumor treatment, which are critical for further experimental validation.

Keywords: drug repositioning, anti-tumor drug, gene expression, drug target, gene interaction network

INTRODUCTION

Drug repositioning is a new way of applying existing therapeutics to new disease indications. Compared with traditional new drug development methods, the advantage of drug repositioning is that it can reduce the time and cost of drug development, and the drug composition has been proven to be safe in human body, so phase I clinical trials can be skipped (1, 2).

The failure probability of new drugs in the development process is about 90% (3), which leads to high drug development costs. In addition, repurposed drugs can save most of the cost of early research and significantly reduce the transition from laboratory research to clinical treatment. According to a research report released by Deloitte & Touche in 2016, according to the tracking results of 12 large pharmaceutical companies for 6 years, the return on investment of R&D giants dropped from 10.1% in 2010 to 3.7% in 2016. It was also calculated that the average cost of developing a new drug has increased from less than 1.2 billion US dollars to 1.54 billion US dollars, and it takes 14 years to launch a new drug (4). Nosengo concluded that it currently takes more than 10 years to bring a drug to the market, and the average research cost is between \$2 billion to \$3 billion. Although the number of approved drugs for development remains the same or decreases over time, the cost of research continues

to increase. In contrast, some studies suggest that repositioning a known drug costs an average of \$300 million, and it takes about six to seven years (5). New solutions are needed to solve the above-mentioned problems in the development of new drugs, including drug repositioning.

Drug repositioning refers to the matching and identification of existing drugs and new indications, and trying to apply newly discovered drugs to the treatment of diseases other than expected diseases (6). In addition, drug repositioning has promoted the development of cancer research (7). Researchers are committed to finding potential drug molecules that can block the exchange of information between cancer cells, and prevent cancer cells from receiving information that promotes their growth and proliferation. At present, *in silico* and activity-based methods are mainly used to determine the feasibility of drug repositioning. *In silico* methods for drug repositioning are affected by drug-to-disease relationships, or the gene expression response of cell lines after treatment. Combining multiple information levels, the relationship network between target and drug can be identified by means of bioinformatics tools and public databases (8, 9). Due to decades of accumulation of structural information between proteins and pharmacophores, the method has gradually become successful. Compared with *in silico* drug repositioning, computerized drug repositioning has become a promising technology with fast speed and low cost (10).

Since the outbreak of Corona Virus Disease 2019 (COVID-19), it spreads rapidly all over the world. There is an urgent need for effective drugs to treat and alleviate the deterioration of this novel Coronavirus (11, 12). Since the development of a new drug is time-consuming and costly, drug reposition is a feasible way to meet this need (13, 14). The treatment of COVID-19 relied on the experience of clinicians (15, 16). So far, some drugs have been proved effective in relieving and improving the symptoms of novel coronavirus pneumonia (17–22). The drugs against the Middle East respiratory syndrome coronavirus (MERS-CoV) and severe acute respiratory syndrome coronavirus (SARS-CoV), such as Lopinavir/ritonavir, have been proved to inhibit many viruses (22, 23). As a nucleoside drug and RNA polymerase (RdRp) inhibitor, remdesivir can inhibit SARS-CoV-2 RdRp, subgenomic mRNA and subviral genomic RNA to block the synthesis of negative chain RNA, thus inhibiting virus replication and antiviral effect (24–26).

In this review, we present the recent progress on *in silico* methods for repositioning drugs and chemical compounds. In particular, we focus on feature-based methods, matrix decomposition-based methods, network-based methods, and reverse transcriptome-based methods. We review the *in silico* popular methods in the four categories separately.

FEATURE-BASED METHODS

In silico methods of drug compounds and repositioning drugs aims to identify the relationship network between target and drug, which is achieved through bioinformatics tools and public databases. Therefore, it needs to ensure high-resolution

structural information, including drugs targets, gene expression profiles, or disease/phenotype information, which usually produce high-dimensional feature datasets. For instance, the Cancer Cell Line Encyclopedia study (27) contains more than 50000 features, representing the mRNA expression and mutational status of thousands of genes. However, the number of available features is significantly greater than the number of training samples. The use of high-dimensional features can lead to overfitting of the model, in fact, only a few features play a key role in the final prediction of drug sensitivity.

Therefore, a feature-based methods are proposed: (1) can prevent over-fitting and improve model performance; (2) can provide a more cost-effective and faster model; (3) can clearly grasp the basic process of generating data. In **Figure 1A**, we visualize the process of the feature-based method. These are important for understanding the relationships between data in the chemical, clinical domains, and biological fields. Therefore, the research of feature-based drugs sensitivity prediction and individualized treatment methods are very necessary. **Table 1** summarizes the feature-based methods used in a large number of studies.

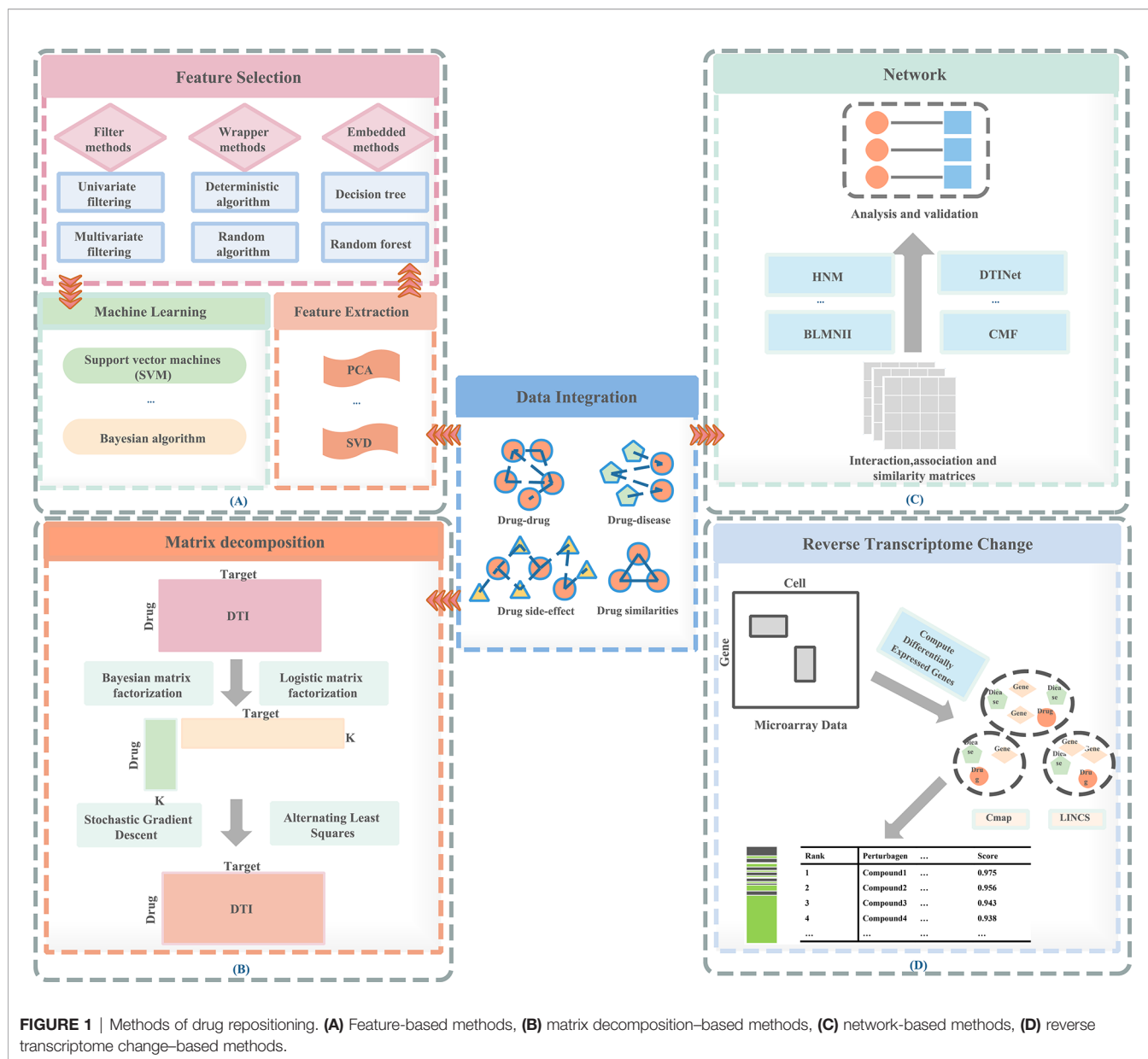
Feature Extraction and Feature Selection

The purpose of feature extraction is to project features into new low-dimensional feature space. The features after dimension reduction are usually a combination of the original features, with the aim of discovering more representative information through the new feature sets. A common example of feature extraction technique is principal component analysis (PCA) (28, 29), which maximizes the variance of each component projection, thereby mapping the original input data to an orthogonal coordinate system.

Feature selection aims to select a small part of the input features without losing the information contained in the original features. Our commonly used feature selection methods include: filter, wrapper and embedded methods.

Filter methods are usually classified according to general features, such as looking at the correlation between individual features or independence and output response. For the prediction of drug sensitivity, our commonly used filtering feature selection methods include: (1) The correlation coefficients between genomic features and output responses (30, 52); (2) ReliefF (31, 32) is general and successful attribute estimators. They are able to detect conditional dependencies between attributes, and provide a unified view of attribute estimation in regression and classification. They have the advantages of low computation cost, robust model and noise tolerant, but cannot distinguish redundant features; and (3) Minimum redundancy maximum relevance (mRMR) (33–35), which reduces the redundancy between features and considers a high degree of statistical dependence and output the response. The advantage of filter methods lies in the low computational cost, which usually leads to the problem of bias, which makes it impossible to determine the multivariate feature relationship.

The quality of the selected features in the wrapper methods is affected by the prediction accuracy of the learning algorithm. The wrapper methods usually use high model accuracy to capture features, but the disadvantage of wrapper methods is that they



overfit the data. Some commonly used wrapper feature selection methods in drug sensitivity prediction include: (1) Sequential floating forward search (SFFS) (36, 37), where in the forward iteration process, the most representative one will select features from the remaining features. If the removed feature has an impact on the improvement of the objective function, it is provided in the floating part; and (2) Recursive feature elimination (38), which is applicable to all feature models, first sorts the features and eliminates the last feature in turn.

The embedded methods select relevant features through the specific structure of the model, which requires the learning process and feature selection to be interrelated. we usually use embedded methods include: Regularization, which penalizes the norm of feature weights, such as ridge regression (39, 40) penalizing the

L-2 norm, LASSO (41, 42, 53) penalize the L-1 norm, and elastic network regularization (43) penalizes the mixture of L-1/2 norm.

In practice, A hybrid methods that combines the most optimal properties of filters and wrappers is usually used. First, the dimension of feature space is reduced by filter methods, and multiple feature subsets can be obtained (44). Then, a wrapper is used to select the optimal feature subset. Several better feature selection methods have been proposed, such as: feature selection based on fuzzy random forest (45), hybrid genetic algorithms (46), hybrid ant colony optimization (47), or hybrid gravity search algorithms (48).

When using hybrid methods, prior knowledge of biological is usually included in the feature section in the process of predicting drug sensitivity. An example is path-based elastic

TABLE 1 | Feature-based methods.

Study	Feature extraction	Feature selection			Description	Ref
		Filter methods	Wrapper methods	Embedded methods		
Pearson	√				An effective algorithm for extracting main feature components of data	(28)
Goswami et al.	√				Use principal components analysis (PCA) to identify and remove abnormal samples	(29)
Costello et al.		√			Bayesian multi-task multiple-kernel learning (MKL) method for drug sensitivity prediction and identification	(30)
Robnik-Sikonja et al.		√			Theory and application of ReliefF and RReliefF	(31)
Haider et al.		√			Multivariate ensemble learning regression tree extension ensemble learning method	(32)
De Jay et al.		√			An extended integration method based on mRMR (mRMRe)	(33)
Peng et al.		√	√		A two-stage feature selection algorithm combining mRMR and other feature selection algorithms	(34)
Liu et al.		√			Minimal redundancy and maximal correlation were used to analyze and predict drug interactions	(35)
Pudil et al.			√		Theory and application of floating search method	(36)
Berlow et al.			√		A sensitivity prediction method based on function perturbation data	(37)
Dong et al.			√		Support vector machine recursive feature elimination	(38)
Tikhonov A				√	Ridge regression penalty L-2 norm	(39)
Neto et al.				√	A Bayesian inference method based on ridge regression	(40)
Tibshirani R				√	LASSO penalizes L-1 norm	(41)
Park et al.				√	RRLASSO method of targeted anticancer drugs	(42)
Zou et al.				√	Regularization of elastic networks based on Mixed Penalties of L-1 and L-2 norms	(43)
Das S		√	√		Mixing of filter and wrapper	(44)
Cadenas et al.		√	√		Feature selection based on fuzzy random forest	(45)
I.S. et al.			√		Hybrid genetic algorithms	(46)
Ali et al.		√	√		Hybrid ant colony optimization	(47)
Sarafrazi et al.			√		Hybrid Gravity Search Algorithm	(48)
Sokolov et al.				√	Path-based elastic network regularization	(49)
Bandyopadhyay et al.				√	A feature selection method combining gene expression data with signals and regulatory pathways	(50)
Amadoz et al.			√		Use the activation state of the signal pathway as a feature	(51)

net regularization (49), which incorporates path knowledge in data-driven feature selection. Feature selection based on biological pathways can select the most important features with minimally redundancy, and combine gene expression data with signaling and regulatory pathways (50) or use the activation state of signaling pathways as features (51).

MATRIX DECOMPOSITION-BASED METHODS

Previously molecular synthesis experiments for drug targets were expensive and time-consuming. Therefore, research on drug repositioning requires effective calculation methods, which have proven to be a viable strategy in the field of *in silico* drug discovery. The basic requirement of calculating drug repositioning is to accurately predict the drug and target (DTIs) interaction. Therefore, researchers have proposed some potential methods for predicting DTI in recent years (Table 2).

We usually use binary labeling matrix Y to represent drug-target interactions (Figure 1). If the drug and the target are in an interaction relationship, it is represented by element 1; If it is not an interactive

relationship, it is represented by 0. The difficulty of predicting DTI lies in whether the known elements in y can accurately predict the labels of unknown elements. To solve these problems, assuming similar drugs tend to similar targets, the similarity between drugs and targets can be used to predict DTI, and vice versa.

Liu et al. proposed a neighborhood regularized logic matrix factorization (NRLMF) method (54). This method uses logical matrix decomposition to simulate the interaction probability of each drug target. We further improve the prediction accuracy by neighborhood regularization. The NRLMF model is the most advanced algorithm and has achieved good results on the basis of five 10-fold cross-validation tests. However, The NRLMF model also has some shortcomings, that is, the drug target interaction information is not considered when the model is established. In response to the above problems, Hao et al. proposed a dual-network integrated logic matrix factorization (DNILMF) (55) and integrated drug target profile information into the model. Based on the NRLMF model, Ban et al. used Gaussian process mutual information to accelerate model parameter search (56). Compared with the previous grid search methods, the method based on Gaussian process mutual information saves about 8.94 times of calculation time. When the area under the curve (AUC) is used for evaluation, the prediction accuracy of the two methods is almost the same.

TABLE 2 | Matrix decomposition-based methods.

Study	Logistic matrix factorization	Bayesian matrix factorization	Probability matrix factorization	Other	Description	Ref
Liu et al.	√				Neighborhood regularized logic matrix factorization (NRLMF)	(54)
Hao et al.	√				Dual network integrated logistic matrix factorization	(55)
Ban et al.	√				The Hyperparameter Optimization of Improved Neighborhood Regularization Logic Matrix Factorization	(56)
Bolgár et al.		√			An extended Bayesian matrix factorization	(57)
Bolgár et al.		√			Variational Bayesian multiple kernel logistic matrix factorization (VB-MK-LMF)	(58)
Gonen		√			A novel Bayesian formula combining matrix factorization and dimensionality reduction	(59)
Peska et al.		√			Matrix decomposition based on Bayesian personalized ranking (BPR)	(60)
Cobanoglu et al.			√		Analyze large-scale interactive networks through probability matrix factorization (PMF)	(61)
Cobanoglu et al.			√		Online tool based on probability matrix factorization method and DrugBank v3	(62)
Zheng et al.				√	Multiple similarities collaborative matrix factorization (MSCMF)	(63)
Wang et al.				√	An improved method of disallow the regular term of the drug pathways	(64)
Ezzat et al.				√	Two matrix factorization methods using graph regularization	(65)
Peng et al.				√	A framework model integrating non-negative matrix factorization, low-rank representation, neighbor interaction profile and sparse representation classification	(34)
Dai et al.				√	A matrix factorization model that combines drugs, diseases and genes with feature vectors of the same dimension	(66)

Bolgár et al. proposed an extended Bayesian matrix factorization method (57), which was combined with a new missing not at random (MNAR) data sub-model. Bolgár et al. later proposed variational Bayesian multiple kernel logistic matrix factorization (VB-MK-LMF) (58), which combines multiple kernel learning, weighted observations and graph Laplacian regularization, and it has explicit modeling probability advantage. Gonen proposed a new Bayesian formula that combines matrix factorization and dimensionality reduction (59). This method uses the chemical similarity of drug components and the genomic similarity of target proteins to predict DTI network. Based on Bayesian personalized ranking (BPR) matrix factorization, Peska et al. proposed a method to predict DTIs (60). They extended BPR by including target deviations, developed a technique for analyzing new drugs, and adjusted the content to take into account the structural similarity between the drug and the target.

Cobanoglu et al. used probabilistic matrix factorization (PMF) to analyze large interaction networks (61). They clustered DrugBank drugs based on PMF latent variables. Cobanoglu et al. later built an online tool for evaluating DTIs (62). They use the PMF method and DrugBank v3, and use the GraphLab collaborative filtering toolkit to train potential variable models.

Zheng et al. proposed a method of multiple similarities collaborative matrix factorization (MSCMF) (63). This method allows the collaborative prediction of DTIs through two low-rank matrices and detects similarities that are important for predicting DTIs. Wang et al. proposed a method to replace the regular term of the drug pathway association matrix (L1 norm) with L2-1 norm (64). Compared with the previous iPad method, this method solves the problem of excessively scattered sparsity, and can obtain more optimized performance by identifying effective drug pathway associations.

Ezzat et al. proposed two matrix factorization methods that use graph regularization and consist of two steps (65). First, convert the binary value in the drug-target matrix Y into an interaction likelihood value. Then use matrix factorization to predict DTI. In cross-validation, it is found that the performance of this method is better than the other three other state-of-the-art methods in most cases. They found that their method reasonably predicted missed interactions with “new drugs” and “new target” simulated cases.

Peng et al. proposed a unified model framework (34), which integrates non-negative matrix factorization, low-rank representation, neighbor interaction profile and sparse representation classification. Dai et al. proposed a matrix factorization model (66), which integrates drugs, diseases and genes with feature vectors of the same dimension. Experiments showed that the integration of genomic space is indeed effective.

NETWORK-BASED METHODS

In the past decade, network-based approaches (**Figure 1**) have been commonly used to predict drug sensitivity (1, 67). We have summarized some network-based methods in **Table 3**. Due to the increase in drug development costs and the decrease in the number of newly approved drugs, it is necessary to determine the new value of existing drugs. Some network-based methods help design unique drug target combinations and combined drugs therapies (68), and improve the treatment of specific patients through powerful channels (69).

Some researchers have proposed that the relationship between drug application, disease treatment, and genes should be studied (70). Some studies analyzed disease diagnosis, treatment, and drug

discovery from the perspective of biological systems and network structure frameworks (71–73). With the development of high-throughput sequencing technology, it is possible to reconstruct cell network and biomolecules. From the cellular level, the reconstructed network will become a hierarchical structure (74). Guney et al. introduced a drug-disease proximity measure that quantifies the interaction between disease and drug targets (84).

Additionally, network-based proximity can help us determine the therapeutic effects of drugs and predict novel drug-disease associations. Kotlyar et al. summarized how drugs disrupt the network, and previous network-based drug effects characterizations included direct binding to partners (75). Drugs can also affect the transcriptome of cells, and networks have been used for the first time to characterize genes differentially regulated by drugs. Cheng et al. constructed a bipartite graph based on the network inference method to predict the interaction between drug and target (76). Chen et al. constructed a general heterogeneous network (77), which was composed of drug and protein, and considered drug-drug chemical similarity, protein-protein sequence similarity and drug-target interaction (78).

The mining potential of drug-disease associations has been consistently used to accelerate the drug repositioning by pharmaceutical companies. Cheng proposed an inference method based on drug-target bipartite network (76), which can be used to predict new targets of known drugs, and described the importance of developing computational methods for predicting potential DTIs. Then, Chen proposed two inference methods, ProbS and HeatS (78), which can predict drug-disease interactions based on the measurement of basic network topology. Methods probs and heats are two methods based on recommendation techniques (79, 80). In order to find the correlation between known drugs and diseases, they solve the above problem by mining the data of drug-disease bipartite network properties. Then, Wang proposed a heterogeneous network model (81). This method uses existing omics data to relocate drugs, diseases and drug targets. This three-layered

heterogeneous network model for drug repositioning captured the interrelationships among diseases, drugs, and targets, with the purpose of novel drug usage prediction. Chen et al. provided a principled method to transfer knowledge from these two domains and improve prediction performance for these two tasks (82). With the help of the relationship between drug target disease, this method urges us to consider drug relocation and drug target prediction in drug discovery.

Some researchers have attempted to reposition drugs by targeting network modules through some unique cases, such as a Parkinson's disease case study. Yue constructed a framework of targeted therapy (83), which combines genome-wide association data with gene co-expression modules of PD disease tissues representing brain regions, and aims to study dysfunctional pathways or processes.

REVERSE TRANSCRIPTOME CHANGE-BASED METHODS

Reverse transcriptome change-based methods (**Figure 1**) are methods based on the gene expression profiles induced by drugs. These methods consider the relationship between drugs, genes, and disease. The publicly accessible gene expression profiles currently include Connectivity Map (CMap, <http://www.broadinstitute.org/cmap>), National Cancer Institute 60 human tumor cell line anticancer drug discovery project (NCI-60 <http://dtp.nci.nih.gov/>), Library of Integrated Cellular Signatures (LINCS <http://www.lincsproject.org/>), and Cancer Cell Line Encyclopedia (CCLE <http://www.broadinstitute.org/ccle>) (2, 52).

It is helpful to facilitate repositioning drugs and chemical compounds with relevant databases. Examples are Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>), The Cancer Genome Atlas (TCGA, <http://tcga-data.nci.nih.gov/tcga/tcgaHome2.jsp>), Gene Expression Database (GXD, [**TABLE 3** | Network-based methods.](http://</p>
</div>
<div data-bbox=)

Study	Description	Ref
Huang et al.	A new system calculation tool called DrugComboRanker prioritizes synergistic drug combinations and reveals its mechanism of action	(68)
Dorel et al.	Drug sensitivity prediction based on high-throughput sequencing data and signal network	(69)
Kanehisa M.	KEGG Mapper tool introduction	(70)
Chen et al.	Alternative techniques and tools for analyzing biomolecular networks	(71)
Zhang A.	Discuss current research problems and solutions in protein-protein interaction networks	(72)
Sun P.G.	A multi-level network model integrating drugs, diseases and genes for disease diagnosis, treatment and drug discovery	(73)
Leiserson et al.	A novel algorithm to find mutated subnetworks (HotNet2) is used	(74)
Guney et al.	A metric for quantifying interactions between drugs, targets, and diseases	(84)
Kotlyar et al.	Use networks to characterize genes that are differentially regulated by drugs and find the differences between the genes regulated by drugs and drug targets	(75)
Cheng et al.	A inference method based on topological similarity of drug target bipartite network	(76)
Chen et al.	A network method based on restart random walk	(77)
Chen et al.	A method based on basic network topology measure is used to predict the direct association between drugs and diseases	(78)
Zhou et al.	A weighting method is used that can be directly applied in extracting hidden network information	(79)
Zhou et al.	Hybrid algorithm based on heat-spreading	(80)
Wang et al.	A computing framework based on heterogeneous network model	(81)
Chen et al.	A principled method to improve the prediction performance of two tasks	(82)
Yue et al.	Reorientation of PD drugs with systemic pharmacology framework	(83)

www.informatics.jax.org/expression.shtml), ArrayExpress (<http://www.ebi.ac.uk/arrayexpress>), et al. The huge amount of publicly available transcriptome data is enabling the repositioning of drugs and chemical compounds based on the gene expression profiles. We summarize the articles based on the above database in **Table 4**.

Lamb et al. established CMAP database (85), which contains more than 6100 gene expression profiles induced by more than 1300 compounds in four cell lines. The main working idea is to enter a query in the CMap database, using the genome of the drug as a reference. Drug candidates with a positive correlation score (the highest is close to 1) may be considered to be related to the reference drug between downstream regulatory and clinical drug response, while drug candidates with a negative correlation score (the lowest is close to -1) may eventually be considered. It is considered that there is no potential correlation or antagonism with the reference drug.

Based on the correlation between drugs and genetic characteristics, we can discover some new drug indications, and assume that drugs with similar characteristics may have similar therapeutic effects (85). Iskar et al. developed a strict filtering and state-of-the-art normalization pipeline for CMap gene expression (86), and it significantly overcomes cross-batch non-biological experimental variation. Hieronymus et al. proposed a chemical genomic method based on gene expression analysis (87), which can be used to discover and predict compounds with cancer phenotypes, for example, for compounds with gedunin and celastrol activity. HSP90 inhibitors are classified. Epoxy anthraquinone derivatives have been found to be a novel DNA topoisomerase inhibitor for the treatment of neuroblastoma and other cancers (88). The alkaloid thaspine from the croton cortex has been shown to play a role in the overexpression of drug efflux transporters in cells, and induce apoptosis of multicellular spheroids cells. It can be used as a dual topoisomerase inhibitor (89).

The molecular mechanism of the traditional Chinese medicinal formula Si-Wu-Tang was discovered through connection maps and gene expression microarray (90). Studies have found that SWT, as an activator and phytoestrogen of Nrf2, it can be used as a non-toxic chemopreventive agent. Through CMap mining and microarray gene expression profiling, the new mechanism of action of traditional Chinese medicine can be verified and discovered. K562 cells exposed to sodium valproate were verified by CMAP database, and it was found that valproate acid could provide certain therapeutic potential in the treatment of leukemia (91). As a combination of approved drugs and failed drugs, repoDB database (<http://apps.chiragjgroup.org/repoDB/>) provides researchers with a simplified hypothesis to prove that all novel predictions are false (92).

In the past, anticancer drugs were screened by transplantable animal tumors. In the late 1980s, NCI-60 cell line dataset was developed by the US National Cancer Institute (NCI), aiming at drug discovery *in vitro* (93). The NCI-60 data set involves nine human cancers with a total of 60 cell lines, including: ovarian cancer, prostate cancer, lung cancer, leukemia, colon cancer, breast cancer, etc. The US National Cancer Institute proposed a comparative algorithm to find new compounds with similar

mechanisms, or possible mechanisms of action of related compounds (94). The similarity search method of bioactivity map can calculate the similarity between drugs according to the bioactivity map of drugs, and relocate the known drugs according to the similarity (95).

Reverse-phase protein lysate microarray is a method for accurately measuring protein expression levels in NCI-60 cell line. This method has a large number of spots and aims to find a type of molecular with high protein/mRNA correlation (96). In February 2016, NCI-60 was no longer supported because NCI decided to use a patient-derived xenograft (PDX) model instead. Since then, some research institutions and drug companies have begun to build their own model PDX library. EurOPDX composed by 16 European institutions jointly consists of 1500 PDX models, The Jackson Laboratory has 450 PDX models, and the drug screening tool released by Novartis uses 1000 PDX models.

The Library of Integrated Network-based Cellular Signatures (LINCS) program was developed by the US National Institutes of Health to increase understanding of normal and diseased cellular states and how to alter them. Researchers at the LINCS transcription center have released a new version of Connectivity Map, which involves 42000 human cells and more than 1.3 million gene expression profiles. This data set is based on L1000 analysis and aims to reduce the cost of gene expression analysis (97).

In order to analyze the effects of different small molecule drugs on six different breast cancer cell lines, the researchers proposed a method to obtain survival measurements and cell growth. Studies have shown that the survival and growth of certain types of breast cancer cells are affected by drugs, and the existence of differences helps to understand the response of breast cancer patients during treatment (98). Studies have shown that the effects of drugs that can reverse the expression of cancer-related genes are beneficial to the treatment of some cancer models (etc. breast, liver, and colon cancer.) (99). They concluded that the four compounds showed high enough potency to reverse gene expression in liver cancer, and used a system-based method to confirm that the four compounds were effective against the discovered liver cancer cell lines.

It is found that the information obtained by different measurement methods under different drug doses has corresponding uniqueness (100), which is conducive to further exploration of drug effects. When researchers examine the variability of drug effects, they need to consider many factors to expand the way they think about drug activity. The conclusion shows that in the comparison of drug reactions, in addition to the drug effect and price, many factors should be considered, such as clinical concentration near and above the IC_{50} .

The Cancer Cell Line Encyclopedia (CCLE) project is an effort to conduct a detailed genetic characterization of a large panel of human cancer cell lines (27). CCLE provides public access analysis and visualization of DNA copy number, mRNA expression, mutation data, and other items for approximately 1000 cancer cell lines, as well as the pharmacological profiles of 24 anti-cancer drugs in 50% of cell lines. Barretina et al. developed the research tools for predicting the genetic variation of cancer drug sensitivity and evaluated their systematic analysis methods. They

TABLE 4 | Reverse transcriptome change-based methods.

Study	CMap	NCI-60	LINCS	CCLE	Description	Ref
Lamb et al.	✓				Establish CMap database	(85)
Iskar et al.	✓				Developed a pipeline for strict filtering and state-of-the-art normalization for gene expression in CMap	(86)
Hieronymus et al.	✓				Regulators for predicting cancer phenotypes based on chemical genomics	(87)
Gheeya et al.	✓				Prediction of the mechanism of action of unknown drugs based on CMap database	(88)
Fayad et al.	✓				Analysis of MCF-7 gene expression in breast cancer cells based on CMap database	(89)
Wen et al.	✓				Detection of gene expression changes caused by traditional Chinese medicine ingredients based on CMAP database	(90)
Zhang et al.	✓				Prediction of molecular mechanism of VPA against CML based on CMAP database	(91)
Brown et al.	✓				A standard database for drug repositioning (repoDB)	(92)
Shoemaker		✓			Review the development and use of the NCI-60	(93)
Zaharevitz et al.		✓			Explain and demonstrate an example of using COMPARE on the web page	(94)
Cheng et al.		✓			Use the NCI-60 data set to identify new targets for drugs and bioactive compounds on a larger scale	(95)
Nishizuka et al.		✓			Proteomic profiling of the NCI-60 cancer cell lines using new high-density reverse-phase lysate microarrays	(96)
Subramanian et al.			✓		Designed a method for cheap and large-scale gene expression analysis (L1000 assay)	(97)
Niepel et al.			✓		Developed a method for cell growth and survival measurements	(98)
Chen et al.			✓		Predicted four highly effective compounds capable of reversing liver cancer gene expression, and confirmed that all four compounds are effective against five liver cancer cell lines	(99)
Fallahi-Sichani et al.			✓		Multi-parameter methods involving analysis	(100)
Barretina et al.				✓	Created a research tool for predicting genetic variation in cancer drug sensitivity	(27)

also applied the prediction model method to the cancer genetic subsets that challenge the current treatment methods.

DISCUSSIONS

We reviewed the four popular *in silico* methods for drug repositioning based on feature, matrix factorization, network, and reverse transcriptome change. Through the analysis of the four methods, we found that each method has its advantages and limitations and more optimal performance can usually be obtained by combining different methods and strategies.

Despite the creation of some excellent drug repositioning models and methods, the development of robust and satisfactory models is still an indispensable process. One of the main problems is the difficulty in developing functional theoretical models or methods, which is challenging because the construction of such models or methods to simulate biological behavior will have a certain degree of complexity. Due to changes in the conditions and environments that exist during different experiments, the gene expression profile may be difficult to define, which results in data discrepancies in gene expression characteristics. In addition, when genes are used as drug targets, gene expression is not always significant, resulting in inaccurate data. Because of these problems, it is difficult for models or methods to identify potential drug target interactions when following chemical structures or molecular mechanisms.

Another major problem associated with the drug repositioning model is the lack of reliable gold standard datasets. In the process of model building, one scheme is to combine the divided training, validation, and test set with k-fold

cross validation and then use the popular evaluation index to evaluate the performance. Another scheme is to establish unique gold standard datasets and then use the evaluation indicators to evaluate the model or method proposed to finally avoid the occurrence of over-fitted problems.

Although there are many challenges in the research of drug repositioning, the integration of multi-source information related to drugs and their side effects, interactions of drugs and diseases, and interactions of drugs and drugs is essential to improve the performance of the drug repositioning domain model. There is still a lack of treatment plans corresponding to the large number of existing diseases, which has inspired more scientific researchers and medical workers to carry out research.

AUTHOR CONTRIBUTIONS

XX and PB conceived the concept of the work. BH, FH, CR and PB performed the experiments. BH, FH, and CR wrote the paper. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by the Natural Science Foundation of China (No. 61803151), the Project of Scientific Research Fund of Hunan Provincial Education Department (Nos. 19A060 and 19C0185), and the Hunan Provincial Innovation Platform and Talents Program (No. 2018RS3105).

REFERENCES

- Yang J, Peng S, Zhang B, Houten S, Schadt E, Zhu J, et al. Human Geroprotector Discovery by Targeting the Converging Subnetworks of Aging and Age-Related Diseases. *Geroscience* (2020) 42(1):353–72. doi: 10.1007/s11357-019-00106-x
- Liu C, Wei D, Xiang J, Ren F, Huang L, Lang J, et al. An Improved Anticancer Drug-Response Prediction Based on an Ensemble Method Integrating Matrix Completion and Ridge Regression. *Mol Ther Nucleic Acids* (2020) 21:676–86. doi: 10.1016/j.omtn.2020.07.003
- DiMasi JA, Hansen RW, Grabowski HG, Lasagna L. Cost of Innovation in the Pharmaceutical Industry. *J Health Econ* (1991) 10(2):107–42. doi: 10.1016/0167-6296(91)90001-4
- Martin L, Hutchens M, Hawkins C, Radnov A. How Much do Clinical Trials Cost? *Nat Rev Drug Discov* (2017) 16(6):381–2. doi: 10.1038/nrd.2017.70
- Nosengo N. New Tricks for Old Drug. *Nature* (2016) 534:314–6. doi: 10.1038/534314a
- Peng L, Tian X, Shen L, Kuang M, Li T, Tian G, et al. Identifying Effective Antiviral Drugs Against SARS-CoV-2 by Drug Repositioning Through Virus-Drug Association Prediction. *Front Genet* (2020) 11:577387. doi: 10.3389/fgene.2020.577387
- Qu Y, Gharbi N, Yuan X, Olsen JR, Blicher P, Dalhus B, et al. Axitinib Blocks Wnt/ β -Catenin Signaling and Directs Asymmetric Cell Division in Cancer. *Proc Natl Acad Sci U S A* (2016) 113(33):9339–44. doi: 10.1073/pnas.1604520113
- Zhou L, Wang J, Liu G, Lu Q, Dong R, Tian G, et al. Probing Antiviral Drugs Against SARS-CoV-2 Through Virus-Drug Association Prediction Based on the KATZ Method. *Genomics* (2020) 112(6):4427–34. doi: 10.1016/j.ygeno.2020.07.044
- Xu X, Long H, Xi B, Ji B, Li Z, Dang Y, et al. Molecular Network-Based Drug Prediction in Thyroid Cancer. *Int J Mol Sci* (2019) 20(2):263. doi: 10.3390/ijms20020263
- Cai Y, Huang T, Yang J. Applications of Bioinformatics and Systems Biology in Precision Medicine and Immunology. *BioMed Res Int* (2018) 2018:1427978. doi: 10.1155/2018/1427978
- Patel A, Jernigan DB. Initial Public Health Response and Interim Clinical Guidance for the 2019 Novel Coronavirus Outbreak - United States, December 31, 2019-February 4, 2020. *MMWR Morb Mortal Wkly Rep* (2020) 69(5):140–6. doi: 10.15585/mmwr.mm6905e1
- Baden LR, Rubin EJ. Covid-19 - The Search for Effective Therapy. *N Engl J Med* (2020) 382(19):1851–2. doi: 10.1056/NEJMe2005477
- Cunningham AC, Goh HP, Koh D. Treatment of COVID-19: Old Tricks for New Challenges. *Crit Care* (2020) 24(1):91. doi: 10.1186/s13054-020-2818-6
- Lu H. Drug Treatment Options for the 2019-New Coronavirus (2019-nCoV). *BioSci Trends* (2020) 14:69–71. doi: 10.5582/bst.2020.01020
- Raoult D, Hsueh P-R, Stefani S, Rolain J-M. COVID-19 Therapeutic and Prevention. (2020) 55:105937. doi: 10.1016/j.ijantimicag.2020.105937
- Wu J, Li W, Shi X, Chen Z, Jiang B, Liu J, et al. Early Antiviral Treatment Contributes to Alleviate the Severity and Improve the Prognosis of Patients With Novel Coronavirus Disease (COVID-19). *J Internal Med* (2020) 288:128–38. doi: 10.1111/joim.13063
- Mitjà O, Clotet B. Use of Antiviral Drugs to Reduce COVID-19 Transmission. *Lancet Global Health* (2020) 8:e639–e640. doi: 10.1016/S2214-109X(20)30114-5
- Pedersen S, Ho Y-C. SARS-CoV-2: A Storm Is Raging. *J Clin Invest* (2020) 130(5):2202–5. doi: 10.1172/JCI137647
- Tang X, Cai L, Meng Y, Xu J, Lu C, Yang J. Indicator Regularized Non-Negative Matrix Factorization Method-Based Drug Repurposing for COVID-19. *Front Immunol* (2021) 11:3824. doi: 10.3389/fimmu.2020.603615
- Ge Y, Tian T, Huang S, Wan F, Li J, Li S, et al. An Integrative Drug Repositioning Framework Discovered a Potential Therapeutic Agent Targeting COVID-19. *Signal Transduct Targeted Ther* (2021) 6(1):165. doi: 10.1038/s41392-021-00568-6
- Jia Z, Song X, Shi J, Wang W, He KL. Transcriptome-Based Drug Repositioning for Coronavirus Disease 2019 (COVID-19). *Pathog Dis* (2020) 78:ftaa036. doi: 10.1093/femspd/ftaa036
- Tang X, Cai L, Meng Y, Xu J, Lu C, Yang J. Indicator Regularized Non-Negative Matrix Factorization Method-Based Drug Repurposing for COVID-19. *Front Immunol* (2020) 11:603615. doi: 10.3389/fimmu.2020.603615
- Mo Y, Fisher D. A Review of Treatment Modalities for Middle East Respiratory Syndrome. *J Antimicrob Chemother* (2016) 71(12):3340–50. doi: 10.1093/jac/dkw338
- Ko W-C, Rolain J-M, Lee N-Y, Chen P, Huang CT, Lee P-I, et al. Arguments in Favor of Remdesivir for Treating SARS-CoV-2 Infections. *Int J Antimicrob Agents* (2020) 55:105933. doi: 10.1016/j.ijantimicag.2020.105933
- Al-Tawfiq J, Al-Homoud A, Memish Z. Remdesivir as a Possible Therapeutic Option for the COVID-19. *Travel Med Infect Dis* (2020) 34:101615. doi: 10.1016/j.tmaid.2020.101615
- Peng L, Shen L, Xu J, Tian X, Liu F, Wang J, et al. Prioritizing Antiviral Drugs Against SARS-CoV-2 by Integrating Viral Complete Genome Sequences and Drug Chemical Structures. *Sci Rep* (2021) 11(1):6248. doi: 10.1038/s41598-021-83737-5
- Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, et al. The Cancer Cell Line Encyclopedia Enables Predictive Modelling of Anticancer Drug Sensitivity. *Nature* (2012) 483(7391):603–7. doi: 10.1038/nature11003
- Pearson KJPM. LIII. On Lines and Planes of Closest Fit to Systems of Points in Space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* (1901) 2: (11):559–72. doi: 10.1080/14786440109462720
- Goswami CP, Cheng L, Alexander PS, Singal A, Li L. A New Drug Combinatory Effect Prediction Algorithm on the Cancer Cell Based on Gene Expression and Dose-Response Curve. *CPT Pharmacometrics Syst Pharmacol* (2015) 4:80–90. doi: 10.1002/psp4.9
- Costello JC, Heiser LM, Georgii E, Gönen M, Menden MP, Wang NJ, et al. A Community Effort to Assess and Improve Drug Sensitivity Prediction Algorithms. *Nat Biotechnol* (2014) 32(12):1202–12. doi: 10.1038/nbt.2877
- Robnik-Šikonja M, Kononenko I. Theoretical and Empirical Analysis of ReliefF and Rrelieff. *Mach Learn* (2003) 53(1):23–69. doi: 10.1023/A:1025667309714
- Haider S, Rahman R, Ghosh S, Pal R. A Copula Based Approach for Design of Multivariate Random Forests for Drug Sensitivity Prediction. *PLoS One* (2015) 10(12):e0144490. doi: 10.1371/journal.pone.0144490
- De Jay N, Papillon-Cavanagh S, Olsen C, El-Hachem N, Bontempi G, Haibe-Kains B. mRMRe: An R Package for Parallelized mRMR Ensemble Feature Selection. *Bioinformatics* (2013) 29(18):2365–8. doi: 10.1093/bioinformatics/btt383
- Peng H, Long F, Ding C. Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy. *IEEE Trans Pattern Anal Mach Intell* (2005) 27(8):1226–38. doi: 10.1109/tpami.2005.159
- Liu L, Chen L, Zhang YH, Wei L, Cheng S, Kong X, et al. Analysis and Prediction of Drug-Drug Interaction by Minimum Redundancy Maximum Relevance and Incremental Feature Selection. *J Biomol Struct Dyn* (2017) 35(2):312–29. doi: 10.1080/07391102.2016.1138142
- Pudil P, Novovicova J, Kittler J. Floating Search Methods in Feature Selection. *Pattern Recognit Lett* (1994) 15:1119–25. doi: 10.1016/0167-8655(94)90127-9
- Berlow N, Davis LE, Cantor EL, Séguin B, Keller C, Pal R. A New Approach for Prediction of Tumor Sensitivity to Targeted Drugs Based on Functional Data. *BMC Bioinformatics* (2013) 14:239. doi: 10.1186/1471-2105-14-239
- Dong Z, Zhang N, Li C, Wang H, Fang Y, Wang J, et al. Anticancer Drug Sensitivity Prediction in Cell Lines From Baseline Gene Expression Through Recursive Feature Selection. *BMC Cancer* (2015) 15:489. doi: 10.1186/s12885-015-1492-6
- Tikhonov ANJO. Solution of Incorrectly Formulated Problems and the Regularization Method. *Soviet Mathematics. Doklady*. (1962) 5.
- Neto EC, Jang IS, Friend SH, Margolin AA. The Stream Algorithm: Computationally Efficient Ridge-Regression Via Bayesian Model Averaging, and Applications to Pharmacogenomic Prediction of Cancer Cell Line Sensitivity. *Pac Symp Biocomput* (2014) 27–38. doi: 10.1142/9789814583220_0004
- Tibshirani R. Regression Shrinkage and Selection Via the LASSO. *J R Stat Soc Series B (Methodol)* (1996) 73(1):273–82. doi: 10.1111/j.1467-9868.2011.00771.x
- Park H, Imoto S, Miyano S. Recursive Random Lasso (RRLasso) for Identifying Anti-Cancer Drug Targets. *PLoS One* (2015) 10(11):e0141869–e. doi: 10.1371/journal.pone.0141869
- Zou H, Hastie T. Regularization and Variable Selection Via the Elastic Net. *J R Stat Soc Series B (Stat Methodol)* (2005) 67(2):301–20. doi: 10.1111/j.1467-9868.2005.00503.x

44. Das S. (2001). Filters, Wrappers and a Boosting-Based Hybrid for Feature Selection. *Proceedings of the Eighteenth International Conference on Machine Learning (Icml-2001)*, pp. 74–81. San Francisco, CA, USA: Morgan Kaufmann.
45. Cadenas JM, Garrido MC, Martínez R. Feature Subset Selection Filter–Wrapper Based on Low Quality Data. *Expert Syst Appl* (2013) 40(16):6241–52. doi: 10.1016/j.eswa.2013.05.051
46. Oh IS, Lee JS, Moon BR. Hybrid Genetic Algorithms for Feature Selection. *IEEE Trans Pattern Anal Mach Intell* (2004) 26(11):1424–37. doi: 10.1109/tpami.2004.105
47. Ali S, Shahzad W. A Feature Subset Selection Method Based on Conditional Mutual Information and Ant Colony Optimization. *Int J Comput Appl* (2012) pp:1–6. doi: 10.1109/ICET.2012.6375420
48. Sarafrazi S, Nezamabadi-pour H. Facing the Classification of Binary Problems With a GSA-SVM Hybrid System. *Math Comput Model* (2013) 57:270–8. doi: 10.1016/j.mcm.2011.06.048
49. Sokolov A, Carlin DE, Paull EO, Baertsch R, Stuart JM. Pathway-Based Genomics Prediction Using Generalized Elastic Net. *PLoS Comput Biol* (2016) 12(3):e1004790. doi: 10.1371/journal.pcbi.1004790
50. Bandyopadhyay N, Kahveci T, Goodison S, Sun Y, Ranka S. Pathway-Based Feature Selection Algorithm for Cancer Microarray Data. *Adv Bioinformatics* (2009) 2009:532989. doi: 10.1155/2009/532989
51. Amadoz A, Sebastian-Leon P, Vidal E, Salavert F, Dopazo J. Using Activation Status of Signaling Pathways as Mechanism-Based Biomarkers to Predict Drug Sensitivity. *Sci Rep* (2015) 5:18494. doi: 10.1038/srep18494
52. Liu X, Yang J, Zhang Y, Fang Y, Wang F, Wang J, et al. A Systematic Study on Drug-Response Associated Genes Using Baseline Gene Expressions of the Cancer Cell Line Encyclopedia. *Sci Rep* (2016) 6:22811. doi: 10.1038/srep22811
53. Yang J, Zhang T, Wan XF. Sequence-Based Antigenic Change Prediction by a Sparse Learning Method Incorporating Co-Evolutionary Information. *PLoS One* (2014) 9(9):e106660. doi: 10.1371/journal.pone.0106660
54. Liu Y, Wu M, Miao C, Zhao P, Li XL. Neighborhood Regularized Logistic Matrix Factorization for Drug-Target Interaction Prediction. *PLoS Comput Biol* (2016) 12(2):e1004760. doi: 10.1371/journal.pcbi.1004760
55. Hao M, Bryant SH, Wang Y. Predicting Drug-Target Interactions by Dual-Network Integrated Logistic Matrix Factorization. *Sci Rep* (2017) 7:40376. doi: 10.1038/srep40376
56. Ban T, Ohue M, Akiyama Y. Efficient Hyperparameter Optimization by Using Bayesian Optimization for Drug-Target Interaction Prediction. 2017 IEEE 7th International Conference on Computational Advances in Bio and Medical Sciences (ICCBS) (2017) pp. 1–6. doi: 10.1109/ICCBS.2017.8114299
57. Bolgár B, Antal P. Bayesian Matrix Factorization With Non-Random Missing Data Using Informative Gaussian Process Priors and Soft Evidences. *J Mach Learn Res* (2016) pp. 25–36.
58. Bolgár B, Antal P. VB-MK-LMF: Fusion of Drugs, Targets and Interactions Using Variational Bayesian Multiple Kernel Logistic Matrix Factorization. *BMC Bioinformatics* (2017) 18(1):440. doi: 10.1186/s12859-017-1845-z
59. Gonen MJB. Predicting Drug-Target Interactions From Chemical and Genomic Kernels Using Bayesian Matrix Factorization. (2012) 28:2304–10. doi: 10.1093/bioinformatics/bts360
60. Peska L, Buza K, Koller J. Drug-Target Interaction Prediction: A Bayesian Ranking Approach. *Comput Methods Programs Biomed* (2017) 152:15–21. doi: 10.1016/j.cmpb.2017.09.003
61. Cobanoglu MC, Liu C, Hu F, Oltvai ZN, Bahar I. Predicting Drug-Target Interactions Using Probabilistic Matrix Factorization. *J Chem Inf Model* (2013) 53(12):3399–409. doi: 10.1021/ci400219z
62. Cobanoglu MC, Oltvai ZN, Taylor DL, Bahar I. BalestraWeb: Efficient Online Evaluation of Drug-Target Interactions. *Bioinformatics* (2015) 31(1):131–3. doi: 10.1093/bioinformatics/btu599
63. Zheng X, Ding H, Mamitsuka H, Zhu S. Collaborative Matrix Factorization With Multiple Similarities for Predicting Drug-Target Interactions. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining ACM. (2013) p. 1025–33. doi: 10.1145/2487575.2487670
64. Wang DQ, Gao YL, Liu JX, Zheng CH, Kong XZ. Identifying Drug-Pathway Association Pairs Based on L1L2,1-Integrative Penalized Matrix Decomposition. *Oncotarget* (2017) 8(29):48075–85. doi: 10.18632/oncotarget.18254
65. Ezzat A, Zhao P, Wu M, Li XL, Kwok CK. Drug-Target Interaction Prediction With Graph Regularized Matrix Factorization. *IEEE/ACM Trans Comput Biol Bioinf* (2017) 14(3):646–56. doi: 10.1109/tcbb.2016.2530062
66. Dai W, Liu X, Gao Y, Chen L, Song J, Chen D, et al. Matrix Factorization-Based Prediction of Novel Drug Indications by Integrating Genomic Space. *Comput Math Methods Med* (2015) 2015:275045. doi: 10.1155/2015/275045
67. Ning K, Zhao X, Poetsch A, Chen WH, Yang J. Computational Molecular Networks and Network Pharmacology. *BioMed Res Int* (2017) 2017:7573904. doi: 10.1155/2017/7573904
68. Huang L, Li F, Sheng J, Xia X, Ma J, Zhan M, et al. DrugComboRanker: Drug Combination Discovery Based on Target Network Analysis. *Bioinformatics* (2014) 30(12):i228–36. doi: 10.1093/bioinformatics/btu278
69. Dorel M, Barillot E, Zinovyev A, Kuperstein I. Network-Based Approaches for Drug Response Prediction and Targeted Therapy Development in Cancer. *Biochem Biophys Res Commun* (2015) 464(2):386–91. doi: 10.1016/j.bbrc.2015.06.094
70. Kanehisa M. Molecular Network Analysis of Diseases and Drugs in KEGG. *Methods Mol Biol (Clifton NJ)* (2013) 939:263–75. doi: 10.1007/978-1-62703-107-3_17
71. Sun PG. The Human Drug-Disease-Gene Network. *Inf Sci* (2015) 306:70–80. doi: 10.1016/j.ins.2015.01.036
72. Zhang A. *Protein Interaction Networks: Computational Analysis*. Cambridge University Press (2009) xiii–xiv doi: 10.1017/CBO9780511626593
73. Chen L, Wang R, Zhang XS. *Biomolecular Networks: Methods and Applications in Systems Biology*. Hoboken, New Jersey: John Wiley & Sons Inc (2009).
74. Leiserson MD, Vandin F, Wu HT, Dobson JR, Eldridge JV, Thomas JL, et al. Pan-Cancer Network Analysis Identifies Combinations of Rare Somatic Mutations Across Pathways and Protein Complexes. *Nat Genet* (2015) 47(2):106–14. doi: 10.1038/ng.3168
75. Kotlyar M, Fortney K, Jurisica I. Network-Based Characterization of Drug-Regulated Genes, Drug Targets, and Toxicity. *Methods (San Diego Calif)* (2012) 57(4):499–507. doi: 10.1016/j.ymeth.2012.06.003
76. Cheng F, Liu C, Jiang J, Lu W, Li W, Liu G, et al. Prediction of Drug-Target Interactions and Drug Repositioning Via Network-Based Inference. *PLoS Comput Biol* (2012) 8(5):e1002503. doi: 10.1371/journal.pcbi.1002503
77. Chen X, Liu MX, Yan GY. Drug-Target Interaction Prediction by Random Walk on the Heterogeneous Network. *Mol Biosyst* (2012) 8(7):1970–8. doi: 10.1039/c2mb00002d
78. Chen H, Zhang H, Zhang Z, Cao Y, Tang W. Network-Based Inference Methods for Drug Repositioning. *Comput Math Methods Med* (2015) 2015:130620. doi: 10.1155/2015/130620
79. Zhou T, Ren J, Medo M, Zhang YC. Bipartite Network Projection and Personal Recommendation. *Phys Rev E Stat Nonlin Soft Matter Phys* (2007) 76(4 Pt 2):46115. doi: 10.1103/PhysRevE.76.046115
80. Zhou T, Kuscsik Z, Liu JG, Medo M, Wakeling JR, Zhang YC. Solving the Apparent Diversity-Accuracy Dilemma of Recommender Systems. *Proc Natl Acad Sci U S A* (2010) 107(10):4511–5. doi: 10.1073/pnas.1000488107
81. Wang W, Yang S, Zhang X, Li J. Drug Repositioning by Integrating Target Information Through a Heterogeneous Network Model. *Bioinformatics* (2014) 30(20):2923–30. doi: 10.1093/bioinformatics/btu403
82. Chen H, Cheng F, Li J. IDrug: Integration of Drug Repositioning and Drug-Target Prediction Via Cross-Network Embedding. *PLoS Comput Biol* (2020) 16:e1008040. doi: 10.1371/journal.pcbi.1008040
83. Yue Z, Arora I, Zhang EY, Laufer V, Bridges SL, Chen JY. Repositioning Drugs by Targeting Network Modules: A Parkinson's Disease Case Study. *BMC Bioinformatics* (2017) 18(Suppl 14):532. doi: 10.1186/s12859-017-1889-0
84. Guney E, Menche J, Vidal M, Barabasi AL. Network-Based in Silico Drug Efficacy Screening. *Nat Commun* (2016) 7:10331. doi: 10.1038/ncomms10331
85. Lamb J. The Connectivity Map: A New Tool for Biomedical Research. *Nat Rev Cancer* (2007) 7(1):54–60. doi: 10.1038/nrc2044
86. Iskar M, Campillos M, Kuhn M, Jensen LJ, van Noort V, Bork P. Drug-Induced Regulation of Target Expression. *PLoS Comput Biol* (2010) 6(9):e1000925. doi: 10.1371/journal.pcbi.1000925
87. Hieronymus H, Lamb J, Ross KN, Peng XP, Clement C, Rodina A, et al. Gene Expression Signature-Based Chemical Genomic Prediction Identifies a Novel Class of HSP90 Pathway Modulators. *Cancer Cell* (2006) 10(4):321–30. doi: 10.1016/j.ccr.2006.09.005
88. Gheeya J, Johansson P, Chen QR, Dexheimer T, Metaferia B, Song YK, et al. Expression Profiling Identifies Epoxy Anthraquinone Derivative as a DNA Topoisomerase Inhibitor. *Cancer Lett* (2010) 293(1):124–31. doi: 10.1016/j.canlet.2010.01.004

89. Fayad W, Fryknäs M, Brnjic S, Olofsson MH, Larsson R, Linder S. Identification of a Novel Topoisomerase Inhibitor Effective in Cells Overexpressing Drug Efflux Transporters. *PLoS One* (2009) 4(10):e7238. doi: 10.1371/journal.pone.0007238
90. Wen Z, Wang Z, Wang S, Ravula R, Yang L, Xu J, et al. Discovery of Molecular Mechanisms of Traditional Chinese Medicinal Formula Si-Wu-Tang Using Gene Expression Microarray and Connectivity Map. *PLoS One* (2011) 6(3):e18278. doi: 10.1371/journal.pone.0018278
91. Zhang XZ, Yin AH, Lin DJ, Zhu XY, Ding Q, Wang CH, et al. Analyzing Gene Expression Profile in K562 Cells Exposed to Sodium Valproate Using Microarray Combined With the Connectivity Map Database. *J Biomed Biotechnol* (2012) 2012:654291. doi: 10.1155/2012/654291
92. Brown AS, Patel CJ. A Standard Database for Drug Repositioning. *Sci Data* (2017) 4:170029. doi: 10.1038/sdata.2017.29
93. Shoemaker RH. The NCI60 Human Tumour Cell Line Anticancer Drug Screen. *Nat Rev Cancer* (2006) 6(10):813–23. doi: 10.1038/nrc1951
94. Zaharevitz DW, Holbeck SL, Bowerman C, Svetlik PA. COMPARE: A Web Accessible Tool for Investigating Mechanisms of Cell Growth Inhibition. *J Mol Graphics Model* (2002) 20(4):297–303. doi: 10.1016/s1093-3263(01)00126-7
95. Cheng T, Li Q, Wang Y, Bryant SH. Identifying Compound-Target Associations by Combining Bioactivity Profile Similarity Search and Public Databases Mining. *J Chem Inf Model* (2011) 51(9):2440–8. doi: 10.1021/ci200192v
96. Nishizuka S, Charboneau L, Young L, Major S, Reinhold WC, Waltham M, et al. Proteomic Profiling of the NCI-60 Cancer Cell Lines Using New High-Density Reverse-Phase Lysate Microarrays. *Proc Natl Acad Sci U S A* (2003) 100(24):14229–34. doi: 10.1073/pnas.2331323100
97. Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, et al. A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* (2017) 171(6):1437–52.e17. doi: 10.1016/j.cell.2017.10.049
98. Niepel M, Hafner M, Duan Q, Wang Z, Paull EO, Chung M, et al. Common and Cell-Type Specific Responses to Anti-Cancer Drugs Revealed by High Throughput Transcript Profiling. *Nat Commun* (2017) 8(1):1186. doi: 10.1038/s41467-017-01383-w
99. Chen B, Ma L, Paik H, Sirota M, Wei W, Chua MS, et al. Reversal of Cancer Gene Expression Correlates With Drug Efficacy and Reveals Therapeutic Targets. *Nat Commun* (2017) 8:16022. doi: 10.1038/ncomms16022
100. Fallahi-Sichani M, Honarnejad S, Heiser LM, Gray JW, Sorger PK. Metrics Other Than Potency Reveal Systematic Variation in Responses to Cancer Drugs. *Nat Chem Biol* (2013) 9(11):708–14. doi: 10.1038/nchembio.1337

Conflict of Interest: CR was employed by Geneis Beijing Co., Ltd.

The remaining 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.

Copyright © 2021 He, Hou, Ren, Bing and Xiao. 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.