Send Orders for Reprints to reprints@benthamscience.net

Current Drug Metabolism, 2019, *20*, 194-202

REVIEW ARTICLE

Recent Advances in the Machine Learning-based Drug-target Interaction Prediction

Wen Zhang[1,*], Weiran Lin[1], Ding Zhang[1], Siman Wang[1], Jingwen Shi[2] and Yanqing Niu[3]

[1]*School of Computer Science, Wuhan University, Wuhan 430072, China;* [2]*School of Mathematics and Statistics, Wuhan University, Wuhan 430072, China;* [3]*School of Mathematics and Statistics, South-Central University for Nationalities, Wuhan 430074, China*

ARTICLE HISTORY

Received: October 16, 2017
Revised: January 18, 2018
Accepted: March 19, 2018

DOI:
10.2174/1389200219666180821094047

Abstract: ***Background*:** The identification of drug-target interactions is a crucial issue in drug discovery. In recent years, researchers have made great efforts on the drug-target interaction predictions, and developed databases, software and computational methods.

***Results*:** In the paper, we review the recent advances in machine learning-based drug-target interaction prediction. First, we briefly introduce the datasets and data, and summarize features for drugs and targets which can be extracted from different data. Since drug-drug similarity and target-target similarity are important for many machine learning prediction models, we introduce how to calculate similarities based on data or features. Different machine learning-based drug-target interaction prediction methods can be proposed by using different features or information. Thus, we summarize, analyze and compare different machine learning-based prediction methods.

***Conclusion*:** This study provides the guide to the development of computational methods for the drug-target interaction prediction.

Keywords: Machine learning, drug-target interaction, drug discovery, drug repurposing, molecular fingerprint, similarity measure.

1. INTRODUCTION

Drugs are chemicals that treat, cure or diagnose human diseases and protect human health. Drug discovery is a time-consuming and extremely expensive industry, and designing a new drug may take 15 years and $800 ~ $1,000 million [1, 2]. Drugs exert their therapeutic effect by changing downstream processes of their targets, and traditional drug discovery approaches follow the paradigm "one molecule-one target-one disease", and tested if a specific protein is a drug target for a specific therapeutic indication. The identification of Drug-target Interactions (DTIs) can find novel targets for existing drugs, namely drug repositioning (repurposing), and can also find new therapeutic effects for pre-approved drugs. Therefore, identifying drug-target interactions is a critical task, which attracts great attention from industries and academies.

Drug targets are cellular proteins from 130 protein families, and four types of targets: enzymes, G-protein-coupled receptors (GPCRs), ion channels and nuclear hormone receptors gain the great attention. The estimated number of drug targets in the human genome is about 6,000-8,000 [3], but only hundreds of targets have been discovered. Wet experiments (*in vitro*) for drug-target interaction identification have the high expense and take a long time. For efficiency and effectiveness, computational (*in silico*) methods are developed to narrow the scope of experimental candidates and provide evidence for their experimental results.

As far as we know, researchers constructed a great number of datasets for drug-target interactions, drug data and target data. Therefore, researchers can develop computational methods to predict drug-target interactions.

The early computational methods are molecular docking. Molecular docking predicts complexes of docking small molecule ligands (drugs) to the appropriate target (proteins) binding sites, by optimizing the energy scoring functions. Docking methods require three-dimensional structures of drugs and target proteins, but structures of most proteins are still unavailable. Due to limitations of docking methods, machine learning techniques are introduced into the drug-target interaction prediction. Machine learning has the capability of finding information in complicated data and allows predictions for large-scale data.

We have searched publications in Web of Science since 2008, by using two keywords: machine learning and drug-target interactions, and obtained 80 publications. Publications per year and their sources are shown in Fig. (**1**). Clearly, the machine learning-based drug-target interaction prediction attracts great attention, and publications have been increasing.

In this paper, we review the recent advances in machine learning-based drug-target interaction prediction from the viewpoint of data science. Section 2 introduces drug-target interaction-related databases, features for drugs and targets, and similarity measures for drugs and targets. Section 3 introduces, discusses and compares machine learning-based computational methods for the drug-target interaction prediction.

2. DATA RESOURCES

Publicly available databases provide data to facilitate the drug-target interaction prediction. In order to predict drug-target interactions, researchers usually collect three types of data: drug-target interactions, drug data and target data, and then extract features for drugs and targets. Moreover, drug-drug similarities and target-target similarities are important for many drug-target interaction prediction methods, and the similarities should be calculated based on features or related data.

2.1. Databases

DrugBank (http://www.drugbank.ca/) is a bioinformatics and cheminformatics resource [4, 5] that includes detailed drug data (*i.e.* chemical, pharmacological and pharmaceutical) with compre-

*Address correspondence to this author at the School of Computer Science, Wuhan University, Wuhan 430072, China; Tel: 86-18986211635 E-mail: zhangwen@whu.edu.cn

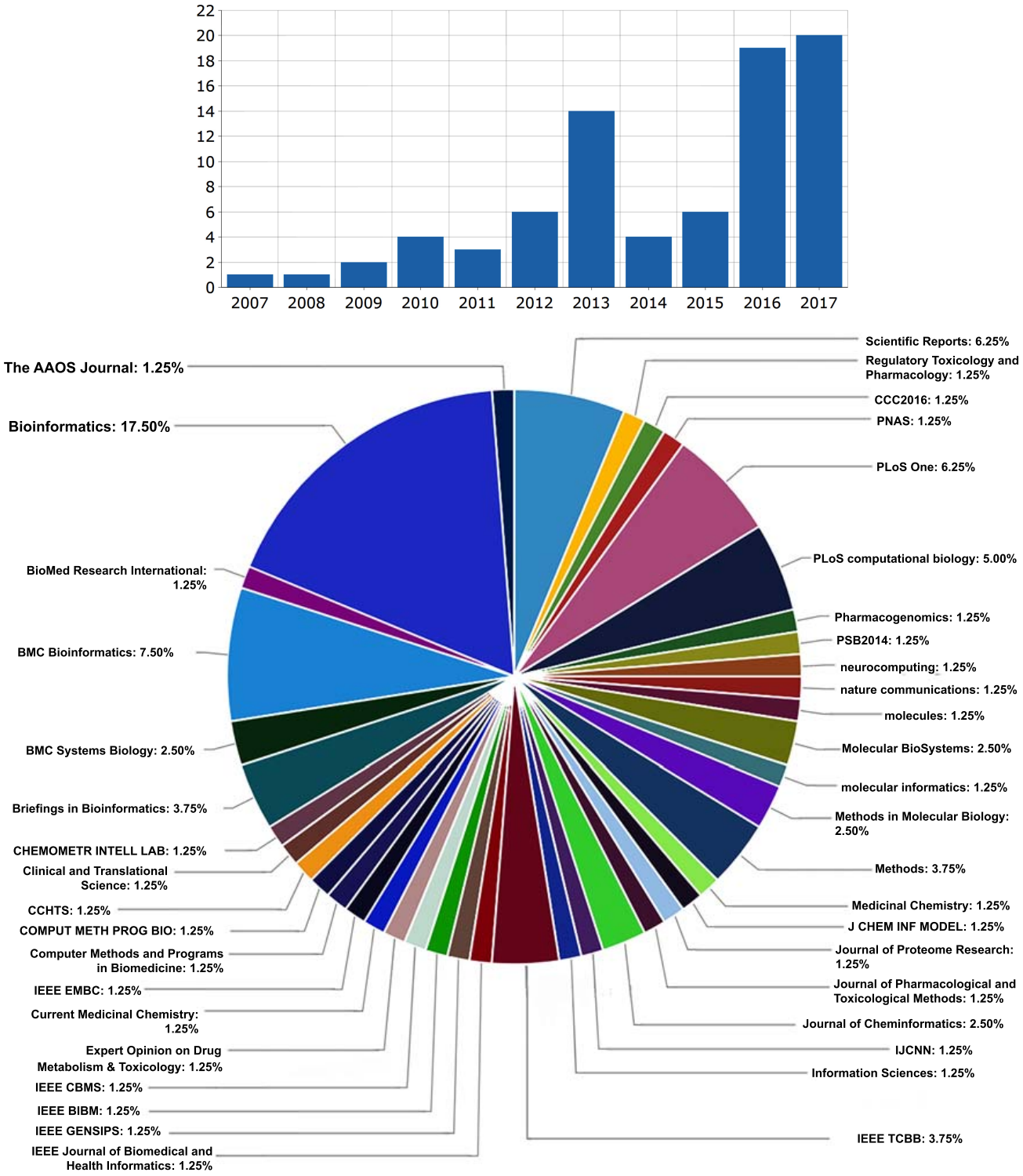

Fig. (1). The numbers of publications per year (left) and the journals for the publications (right).

hensive drug target data (*i.e.* sequence, structure and pathway). The database has 9,591 drug entries, and drug entries are annotated by 4,661 non-redundant proteins (*i.e.* drug target/enzyme/transporter/carrier). Therefore, researchers can obtain drug-target interactions as well as other data describing drugs and targets.

KEGG DRUG (http://www.genome.jp/kegg/drug/) is a comprehensive drug information resource for approved drugs in Japan, USA, and Europe unified based on the chemical structures and/or the chemical components, and associated with the therapeutic target, metabolizing enzyme, and other molecular interaction network information [6, 7]. KEGG DRUG contains drug target molecules in the context of KEGG pathway maps. Drug metabolizing enzymes and transporters are provided, and adverse drug-drug interaction data were collected from the Japanese package inserts. In addition, therapeutic category, ATC code and classification of drugs are described in the databases.

ChEMBL (https://www.ebi.ac.uk/chembl/) is a database of bioactive drug-like small molecules [8]. ChEMBL collected data from the scientific literature. ChEMBL database contains compound bioactivity data against drug targets, and also provides visualized tools for the data analysis.

STITCH (http://stitch.embl.de/) is a database of known and predicted interactions between chemicals and proteins [9]. Interactions in STITCH are collected from five sources: genomic context predictions, high-throughput lab experiments, co-expression, automated text-mining and previous knowledge in databases.

MATADOR (http://matador.embl.de/) is a resource for protein-chemical interactions [10]. The database contains binding and interactions between proteins and chemicals from automated text-mining and manual annotation.

SUPERTARGET(http://insilico.charite.de/supertarget/) is a database for drug-target relations [11]. The database has three different types of entities: drugs, proteins, side effects, and these entities are connected between each other through drug-protein, protein-protein and drug-side-effect relations. Moreover, target-target interactions and sequence similarities between the targets are involved; proteins are retrieved from UniProt and are displayed with synonyms and organism information; ATC-codes, structure information, binding affinities and side-effects are annotated for drugs.

Therapeutic Target Database (TTD) (http://bidd.nus.edu.sg/group/cjttd/) [12] contains 2,589 targets (successful targets, clinical trial targets, and research targets), and 31,614 drugs (approved drugs, clinical trial drugs, experimental drugs). In entries, functions, sequences, 3D structures, ligand binding properties and enzyme nomenclatures are annotated for targets; structures, therapeutic classes and clinical development status are annotated for drugs.

BindingDB (http://www.bindingdb.org/bind/index.jsp) is a public, web-accessible database [13]. BindingDB contains 1,391,403 binding data, for 7,225 protein targets and 621,060 small molecules. There are 2,291 protein-ligand crystal structures with BindingDB affinity measurements for proteins with 100% sequence identity, and 5,816 crystal structures allowing proteins to 85% sequence identity.

BRENDA (http://www.brenda-enzymes.org/) is an information system [14] with enzyme-specific data manually extracted from over 140,000 different scientific articles. Each enzyme entry is clearly linked to at least one literature reference, to its source organism, and the protein sequence of the enzyme.

PubChem Compound (https://pubchem.ncbi.nlm.nih. gov/) is a database of chemical molecules and their activities against biological assays [15]. The database is maintained by the National Center for Biotechnology Information (NCBI). Millions of compound structures and descriptive datasets can be freely downloaded *via* FTP.

UniProt(http://www.uniprot.org/) is a freely accessible database of protein sequence and functional information [16], which are derived from literature. UniProt provides four components: UniProtKB/Swiss-Prot, UniParc, UniRef, and UniMes. In addition, UniProtKB/Swiss-Prot is a manually annotated, non-redundant protein sequence database.

BioGRID (https://thebiogrid.org/) is a database of protein-protein interactions, genetic interactions, chemical interactions, and post-translational modifications [17].

SIDER (http://sideeffects.embl.de/) contains information on marketed medicines and their adverse drug reactions [18], which are extracted from public documents and package inserts. The side effect frequency, drug and side effect classifications as well as links to further information are also provided.

GO database (http://geneontology.org/page/go database) is a relational database comprised of the GO ontologies as well as the annotations of genes and gene products to terms in those ontologies [19].

Pfam (http://pfam.xfam.org/) is a database containing more than 13 000 manually curated protein families [20]. More importantly, all detectable protein sequences and their family information are provided.

The International Classification of Disease (ICD) (http://www.who.int/classifications/icd/en/) is designed as a health care classification system, providing a system of diagnostic codes for classifying diseases.

2.2. Data for Drug-Target Interaction Prediction

Data can be obtained from publicly available databases for the drug-target interaction prediction.

Known drug-target interactions should be provided no matter what machine learning methods are adopted. The known drug-target interactions can be obtained from the DrugBank database, KEGG DRUG, BRENDA, SuperTarget, and etc. Moreover, data which describe the characteristics of drugs and targets can be obtained from databases. The data available in databases are shown in Fig. (**2**). The rows represent databases; the columns are data; a grey entry means that the data are provided in the database.

2.2.1. Data for Drugs

Drug chemical structures can be considered to be one of the most important features, for structures play the most important roles in exerting functions.

Drug pathways are usually considered as the important factor for drug functions. In biochemistry, a metabolic pathway is a linked series of chemical reactions occurring within a cell. Pathways reflect the effects of drugs, and maybe influence drug-target interactions.

Drug-drug interactions can be the indicator for drug-target interactions. Drug-drug interactions occur when a drug affects the activity of another drug, and the drug-drug interaction information may be useful for identifying drug-target interactions.

Drug transporters may affect drug-target interactions. In cellular biology, membrane transport refers to the collection of mechanisms that regulate the passage of solutes such as ions and small molecules through biological membranes, which are lipid bilayers that contain proteins embedded in them.

Drug carriers influence the drug-target interactions. A drug carrier is any substrate used in the process of drug delivery which serves to improve the selectivity, effectiveness, and safety of drug administration.

ATC codes indicate the effects of drugs and may be related to drug-target interactions. ATC system is the Anatomical Therapeutic Chemical Classification System. The classification system divides drugs into different groups, according to the organ or system on which they act their therapeutic and chemical characteristics.

Drug side effects are widely used for the drug-target interaction prediction. Drug interacts with targets to exert effects. In medicine, a side effect is the unintended effects (therapeutic or adverse), and all drugs have side effects.

2.2.2. Data for Targets

Drug targets are cellular proteins, and target protein sequences are the most important data for targets. Protein sequences, also known as primary structures, are the linear sequence of amino acids in a peptide or protein.

Gene Ontology (GO) defines concepts and classes to describe gene function and relationships between these concepts. GO classifies functions along three aspects: molecular function, cellular component and biological process. GO data can be downloaded from GO database.

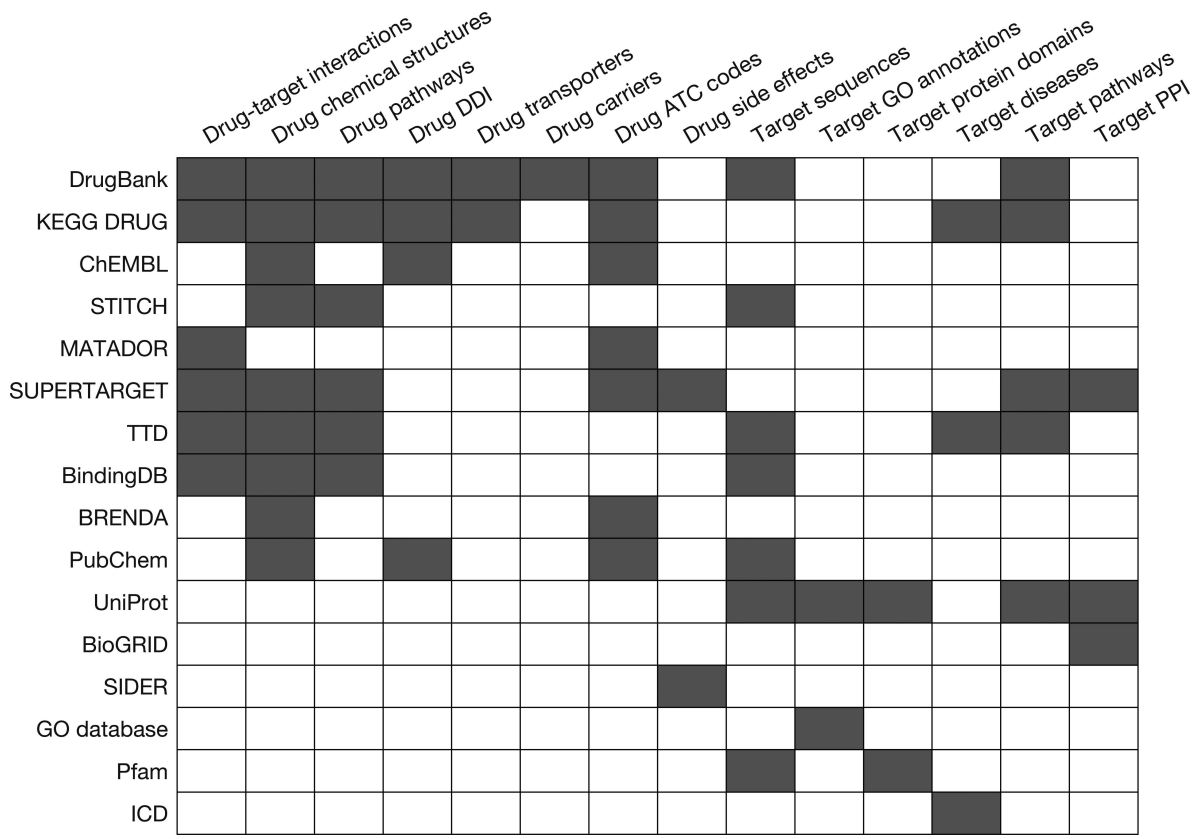

Fig. (2). Data available in databases.

Table 1. Software/Server and its calculated fingerprints.

Software/Server	#	Types of Fingerprints
CDK	16	ECPF; FCFP; Daylight-type fingerprints; MACCS fingerprints; Estate fingerprints; Extended fingerprints; Pubchem fingerprints; Hybridization fingerprints; KlekotaRoth fingerprints; Lingo fingerprints; GraphOnly fingerprints; Substructure fingerprints; Signature fingerprints; IntArray fingerprints; IntArrayCount fingerprints; BitSet fingerprints
PaDEL	12	Fingerprints; MACCS fingerprints; Estate fingerprints; Extented fingerprints; Pubchem fingerprints; KlekotaRoth fingerprints; KlekotaRoth fingerprintsCount; GraphOnly fingerprints; Substructure fingerprints; Substructure fingerprintsCount; AtomPairs2D fingerprints; AtomPairs2D fingerprintsCount
RDKit	8	RDK fingerprints; Topological Torsions fingerprints; Atom Pairs fingerprints; Morgan/Circular fingerprints; MACCS keys fingerprints;2D Pharmacophore fingerprints; Pattern fingerprints; Extended Reduced Graphs
Openbabel	4	FP2; FP3; FP4; MACCS fingerprints
JCompoundMapper	19	DFS; ASP; AP2D; AT2D; AP3D; AT3D; CATS2D; CATs3D; PHAP2POINT2D; PHAP3POINT2D; PHAP2POINT3D; PHAP3POINT3D; ECFP; ECFPVariant; LSTAR; SHED; RAD2D; RAD3D; MACCS
ChemoPy	7	Daylight-type fingerprints; FP4; MACCS; E-state; Atom Paris; Torsions; Morgan
ChemDes	42	CDK; PaDEL; RDKit; Openbabel; jCompoundMapper; ChemoPy

Protein domains reflect the classification information or functional information of target proteins. Protein domains for targets can be extracted from Pfam database.

Targets-disease data can be downloaded from the Therapeutic Targets Database. Each target protein is mapped to one or more diseases with International Classification of Disease (ICD) identifiers.

A substantial number of multi-target drugs produce their therapeutic effects *via* activities against multiple targets of the same pathway, and many drug combinations achieve synergistic therapeutic effects by modulating multiple targets in the same pathway. Pathways of targets with KEGG identifiers were obtained from TTD.

Protein-protein Interactions (PPIs) are the physical contacts of high specificity established between two or more protein molecules as a result of biochemical events steered by electrostatic forces including the hydrophobic effect. PPI data can be downloaded from BioGRID.

2.3. Feature Extraction

After obtaining data from databases, features should be extracted from data for modelling. Data, such as transporters, pathways, drug-drug interactions, side effects, *etc*., are descriptive data. In a descriptive data, drugs are annotated by a set of descriptors. For

Table 2. Servers and calculated sequence-derived features.

Server	Features
PROFEAT	Amino acid composition; Dipeptide composition; Normalized Moreau–Broto autocorrelation; Moran autocorrelation; Geary autocorrelation; Composition; Transition; Distribution; Sequence-order-coupling number; Quasi-sequence-order descriptors
Pse-in-One	Basic k-mer; Auto covariance; Cross covariance; Auto-cross covariance; Parallel correlation pseudo amino acid composition; Series correlation pseudo amino acid composition; General parallel correlation pseudo amino acid composition; General series correlation pseudo amino acid composition

example, a drug may have several side effects. Since a descriptive data is actually a set of descriptors, a drug can be described by a subset of descriptors in the feature, and thus represented as a binary feature vector, whose dimension means the presence or absence of descriptors by using the value 1 or 0. Similarly, targets can be represented as feature vectors. For example, each protein was represented by a domain fingerprint (binary vector) whose elements encode the presence or absence of each Pfam domain by 1 or 0.

Moreover, several extraction approaches have been especially designed for drug structures and target protein sequences.

2.3.1. Drug Molecular Fingerprints

Structures of drugs can be described by different formats: SMILES, SDF, InChI, Mol2, CML, *etc*. Researchers can calculate substructure descriptors, namely "molecular fingerprint", from drug structures. As far as we know, several state-of-the-art tools are publicly available.

The Chemistry Development Kit (CDK) (https://cdk.github.io/) is a collection of modular Java libraries for processing chemical information [21]. The software should be installed under Linux. CDK can calculate 16 types of molecular fingerprints.

PaDEL (http://www.yapcwsoft.com/dd/padel descriptor/) is a software to calculate molecular descriptors and fingerprints. The software currently calculates 12 types of fingerprints [22] (Table **1**).

RDKit (http://www.rdkit.org) is a collection of cheminformatics and machine-learning software written in C++ and Python. This software can calculate 8 molecular fingerprints. The software can be used under all operating systems: Linux, OS X and windows, and support python 2 and python 3.

Openbabel (http://openbabel.org) is a chemical toolbox [23]. Openbabel is written in C++, and Openbabel can run on Windows, Linux and MacOSX. Openbabel can calculate 4 types of molecular fingerprints.

jCompoundMapper (http://jcompoundmapper. sourceforge.net) is an open source Java library and command-line tool for chemical fingerprints [24]. It provides popular fingerprinting algorithms for 19 types of molecular fingerprints.

ChemoPy (http://code.google.com/p/pychem/ downloads/list) is a freely available python package for computational biology and cheminformatics [25]. It provides users with comprehensive implementations of 7 types of molecular fingerprints.

ChemDes (http://www.scbdd.com/chemdes/list-fingerprints/) is a free web-based platform for the calculation of molecular descriptors and fingerprints [26], which provides more than 3,679 molecular descriptors from several open source packages: CDK, PaDEL, RDKit, Openbabel, jCompoundMapper and ChemoPy.

2.3.2. Feature Extraction From Target Sequences

For the target protein sequences, there are a great number of encoding approaches, which can transform sequences as feature vectors [27-29]. For example, Composition, Transition and Distribution (CTD) are extensively used sequence-derived physicochemical features. The web server PROFEAT and the web server Pse-in-One can easily generate feature vectors from input sequences, according to the encoding schemes selected by users. These tools have been widely used in bioinformatics [30-34].

PROFEAT (http://jing.cz3.nus.edu.sg/cgi-bin/prof/prof.cgi) computes 6 feature groups composed of ten feature encoding schemes with 51 descriptors and 1,447 descriptor values [35]. The amino acid composition, three autocorrelation features, the composition, transition, distribution features and the sequence-order features are computed with the fraction of amino acid types, the distribution of amino acid properties, the amino acid distribution patterns and the physicochemical distance matrixes of protein sequences, respectively.

Inspired by PseKNC-General [36], Pse-in-One (http:// bioinformatics.hitsz.edu.cn/Pse-in-One/) was proposed, which can generate different feature vectors of DNA, RNA and protein sequences [37]. Pse-in-One contains 8 feature encoding schemes for protein sequences. Compared with the PROFEAT, Pse-in-One is aimed to generate various modes of pseudo-components of protein sequences. By integrating the physicochemical properties of constituent amino acids and higher level information of protein sequences such as functional domain and gene ontology, Pse-in-One encodes protein sequence order information into feature vectors (Table **2**).

2.4. Similarity Measures

How to measure drug-drug similarity or target-target similarity is critical for many drug-target interaction prediction methods. Here, we introduce several schemes of calculating similarities for drugs and targets.

If we can represent drugs (targets) as feature vectors, we can calculate similarities for drugs (targets) based on feature vectors. As far as we know, three popular similarity measures, namely Jaccard similarity, cosine similarity and Gauss similarity, can be adopted to calculate similarity. A novel similarity measure named "linear neighborhood similarity" [38-40] and network-based similarities [41] have been used in the related studies. Moreover, there are still specified approaches for calculating similarities.

The web server SIMCOMP (http://www.genome.jp/ tools/simcomp/) can compute the chemical structure similarities between compounds [42], which provides a global similarity score based on the size of the common substructures between two compounds using a graph alignment algorithm.

The sequence similarities between the proteins can be computed by using the Smith-Waterman algorithm. The Smith-Waterman algorithm performs local sequence alignment, that is, for determining similar regions between two strings of nucleic acid sequences or protein sequences. Instead of looking at the entire sequence, the Smith-Waterman algorithm compares segments of all possible lengths and optimizes the similarity measure. There are other approaches for calculating sequence similarities [43-45].

The semantic similarity score between each pair of proteins is calculated based on the overlap of the GO terms that are associated with the two proteins. All three types of ontologies are used in the computation as similar drugs are expected to interact with proteins that act in similar biological processes or have similar molecular

functions or reside in similar compartments. Jaccard score with respect to the GO terms of each pair of proteins is calculated as their similarity.

3. COMPUTATIONAL METHODS

Recently, a great number of machine learning-based computational methods have been proposed for the drug-target interaction prediction. Here, we analyze 80 publications, mentioned in the introduction. As shown in Fig. (**3**), keywords are visualized in form of the word cloud according to their frequencies. "Similarity" is the word of most frequency. Actually, the principle that similar drugs (or similar targets) may interact with the same target (drug) is widely used to develop similarity-based methods. "Network-based" is another important keyword. Many methods predict interactions based on the drug-target interaction networks, drug-drug similarity-based networks, target-target similarity-based networks or heterogeneous networks. Computational methods are usually associated with a keyword "repositioning", which demonstrates that one important goal of drug-target interaction prediction is drug repositioning.

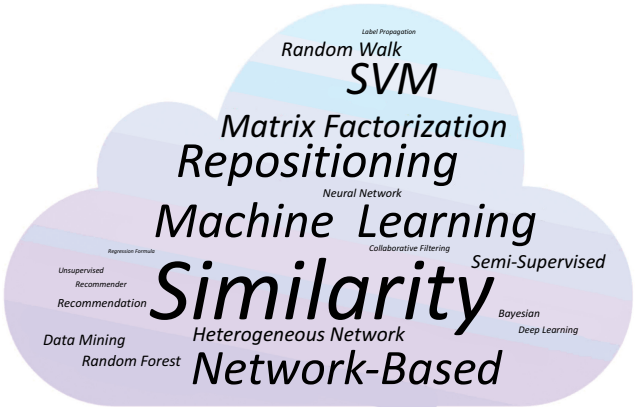

Fig. (3). Visualizing key words of publications in form of a word cloud.

Many machine learning-based drug-target interaction prediction methods are the hybrid of machine learning techniques. For example, similarities are important for the drug-target interaction prediction, and the similarity-based network can be constructed. Therefore, these methods can be considered as the similarity-based methods or network-based methods. Here, we roughly divide the prediction methods into binary classification methods, matrix factorization methods, network inference methods and other methods.

3.1. Binary Classification Methods

Binary classification methods utilize drug features or target features to describe drug-target interaction pairs and non-interaction pairs, and then build binary classification models by using Support Vector Machine (SVM), neutral network, nearest neighbor, *etc*.

Bleakley *et al* [46] extracted features for drugs (or proteins) and built SVM-based bipartite local models to predict drugs (or proteins) interacting with each protein (or drugs). Yamanishi *et al*. [47] proposed the framework of supervised bipartite graph inference to predict unknown drug-target interactions from chemical, genomic and pharmacological data. Tabei *et al*. [48] represented a compound-protein pair by a feature vector corresponds to a pair of a chemical substructure and a protein domain, and utilized logistic regression and SVM to build prediction models. Mei [49] improved the bipartite local model by considering new drug candidates through its neighbors' interaction profiles. Liu *et al*. [50] built up highly credible negative samples for the compound-protein interaction. Mousavian *et al* [51] used the bi-gram features extracted from the Position Specific Scoring Matrix (PSSM) to describe positive instances and negative instances and built SVM-based prediction models. Ezzat *et al*. [52] proposed an ensemble learning method that incorporates techniques to address the issues of between-class imbalance and within-class imbalance. Ding *et al*. [53] used molecular substructure fingerprints, multivariate mutual information (MMI) of proteins and network topology to represent drugs, targets and relationships between them, and then employed SVM and feature selection to construct a model for predicting DTIs. Keum *et al*. [54] categorized unlabeled interactions and negative interactions among unknown interactions using a clustering method, and then adopted a self-training SVM to construct prediction models. Li *et al*. [55] extracted features from drug chemical structures and protein sequences to describe interactions and non-interactions, and then build prediction models by using discriminative vector machine and SVM. Meng *et al* [56] combine Bi-gram probabilities (BIGP) and Position Specific Scoring Matrix (PSSM) to build prediction models by using Relevance Vector Machine (RVM). Peng *et al*. [57] studied how to select negative samples for the drug-target interaction prediction.

Binary classification methods have several limitations. First, these methods need positive instances and negatives instances. Non-interacting drug-target pairs are usually used as negative instances, but it will bring noise and influence the performances. Second, negative instances are much more than positive instances, and data imbalance is a serious problem.

3.2. Network Inference Methods

Network inference methods construct networks (or bipartite) from drug-drug similarities, target-target similarities or drug-target interactions, and then predict interactions based on the network.

Chen *et al*. [58] applied the random walk technique to walk on drug-drug similarity network, target-target similarity network and known drug-target interaction network for predictions. Cheng *et al*. [59] adopted the resource allocation method to infer interactions in the drug-target bipartite network. Alaimo *et al*. [60] extended a well-established recommendation technique by domain-based knowledge including drug and target similarity. Emig *et al* [61] proposed the network-based consensus method. Re *et al*. [62] presented a general framework based on bipartite network projections by which homogeneous pharmacological networks can be constructed and integrated from heterogeneous and complementary sources of chemical, biomolecular and clinical information. Yu *et al*. [63] proposed a new graph-based semi-supervised learning method. Alaimo *et al*. [64] proposed a web-based interface to the DT-Hybrid algorithm, which applies a recommendation technique based on bipartite network projection implementing resources transfer within the network. Seal *et al*. [65] applied random walk with restart (RWR) method to a heterogeneous network of drugs and targets compiled from DrugBank database. Yan *et al*. [66] presented a new method of network-based label propagation with mutual interaction information derived from heterogeneous networks. Fu *et al*. [67] constructed prediction models based on meta-path topological features of an enriched semantic network. Li *et al*. [68] constructed a biological relevant interactome network. Lu *et al*. [69] utilized the missing link prediction method for the drug-target interaction prediction. Luo *et al*. [70] developed a computational pipeline, called DTINet, to predict novel drug-target interactions from a constructed heterogeneous network, which integrates diverse drug-related information. Wu *et al*. [71] developed a useful tool, namely substructure-drug-target network-based inference method, to prioritize potential targets for old drugs, failed drugs and new chemical entities.

Network inference methods don't require negative instances, and usually, make use of known interactions to make predictions.

3.3. Matrix Factorization Methods

In machine learning, the matrix factorization can decompose a data matrix into a product of matrices, and then discover the unknown associations. By formulating drug-target interactions as a matrix, many methods have been developed based on the matrix factorization technique.

Gonen [72] proposed a novel Bayesian formulation that combines dimensionality reduction, matrix factorization and binary classification for predicting drug-target interaction. Zheng *et al*. [73] proposed the collaborative matrix factorization with multiple similarities for predicting drug-target interactions. Liu *et al*. [74] proposed the neighborhood regularized logistic matrix Factorization. The method focuses on modeling the probability that a drug would interact with a target by logistic matrix factorization, where the properties of drugs and targets are represented by drug-specific and target-specific latent vectors, respectively. Hao *et al*. [75] proposed the dual-network integrated logistic matrix factorization. Ezzat *et al*. [76] proposed the graph-regularized matrix factorization method. Peska *et al*. [77] presented the Bayesian personalized ranking matrix factorization method.

3.4. Other Methods

In recent years, deep learning has been used for the drug-target interaction prediction. Hu *et al*. [78] proposed a multi-scale deep representation of features, and then built SVM-based prediction models based on the deep representation. Yuan *et al*. [79] proposed the ranking method which integrates diverse information to predict drug-target interaction for new candidate drugs or targets. Tian *et al*. [80] proposed a deep learning method for compound-protein interactions prediction, which employs a deep neural network to effectively learn the representations of compound-protein pairs. Wen *et al*. [81] developed a deep learning-based algorithmic framework. The method extracted representations from raw input descriptors using unsupervised pretraining and then applies known label pairs of interaction to build a classification model. Zong *et al*. [82] adopted a deep learning method to calculate the similarities within linked tripartite network which is generated from biomedical linked datasets and then developed a similarity-based prediction method.

Moreover, there are many kernel-based methods, such as pair kernel method [83], net Laplacian regularized least squares [84], regularized least squares with Kronecker product kernel [85], Spectral kernel learning [86] and multiple kernel learning [87].

There are still other interesting works. Wang *et al*. [88] utilized restricted Boltzmann machine to predict drug-target interactions. Koohi *et al*. [89] used collaborative filtering in recommender systems for prediction. Fakhraei *et al*. [90] made predictions using probabilistic soft logic. Zhang *et al*. [91] proposed a multi-view DTI prediction method based on clustering. Zhang *et al*. [92] proposed a label propagation method with linear neighborhood information for predicting unobserved drug-target interactions.

CONCLUSION

In this paper, we review the recent advances in the machine learning-based drug-target interaction prediction. Currently, there are lots of databases, which provide diverse information on drugs and targets. In most work, the data are transformed into feature vectors or similarities, and then suitable machine learning methods are adopted to build prediction models. We detailedly summarize the process of data collection, feature representation, similarity calculation and machine learning modeling.

However, there are still bottlenecks for the drug-target interaction prediction. First, interaction data, drug data and target data are stored in different databases and are usually annotated with different IDs, labels or names. Second, known drug-target interactions are limited, and the data become even less if considering features for drugs and targets. Therefore, databases with integrated information and high-quality datasets are keys to the further improvements in the drug-target interaction prediction.

CONSENT FOR PUBLICATION

Not applicable.

FUNDING

This work is supported by the National Natural Science Foundation of China (61772381, 61572368), Fundamental Research Funds for the Central Universities (2042017kf0219).

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

Declared none.

REFERENCES

[1] Breckenridge, A.M. Clinical pharmacology and therapeutics. *BMJ*, **1995**, *310*, 377-380.

[2] Adams, C.; Brantner, V.V. Estimating the cost of new drug development: Is it really $802 million?: Variations in cost estimates suggest that policymakers should not use a single number to characterize drug costs. *Health Aff.,* **2006**, *25*, 420-428.

[3] Russ, A.; Lampel, S. The druggable genome: An update. *Drug Discov. Today*, **2005**, *10*, 1607-1610.

[4] Knox, C.; Law, V.; Jewison, T.; Liu, P.; Ly, S.; Frolkis, A.; Pon, A.; Banco, K.; Mak, C.; Neveu, V.; Djoumbou, Y.; Eisner, R.; Guo, A.C.; Wishart, D.S. DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res.,* **2011**, *39*, D1035-1041.

[5] Law, V.; Knox, C.; Djoumbou, Y.; Jewison, T.; Guo, A.C.; Liu, Y.; Maciejewski, A.; Arndt, D.; Wilson, M.; Neveu, V.; Tang, A.; Gabriel, G.; Ly, C.; Adamjee, S.; Dame, Z.T.; Han, B.; Zhou, Y.; Wishart, D.S. DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res.,* **2014**, *42*, D1091-1097.

[6] Kanehisa, M.; Goto, S.; Furumichi, M.; Tanabe, M.; Hirakawa, M. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.,* **2009**, *38*, D355-D360.

[7] Kanehisa, M.; Furumichi, M.; Tanabe, M.; Sato, Y.; Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.,* **2017**, *45*, D353-D361.

[8] Gaulton, A.; Hersey, A.; Nowotka, M.; Bento, A.P.; Chambers, J.; Mendez, D.; Mutowo, P.; Atkinson, F.; Bellis, L.J.; Cibrian-Uhalte, E.; Davies, M.; Dedman, N.; Karlsson, A.; Magarinos, M.P.; Overington, J.P.; Papadatos, G.; Smit, I.; Leach, A.R. The ChEMBL database in 2017. *Nucleic Acids Res.,* **2017**, *45*, D945-D954.

[9] Szklarczyk, D.; Santos, A.; von Mering, C.; Jensen, L.J.; Bork, P.; Kuhn, M. STITCH 5: augmenting protein-chemical interaction networks with tissue and affinity data. *Nucleic Acids Res.*, **2016**, *44*, D380-384.

[10] Gunther, S.; Kuhn, M.; Dunkel, M.; Campillos, M.; Senger, C.; Petsalaki, E.; Ahmed, J.; Urdiales, E.G.; Gewiess, A.; Jensen, L.J.; Schneider, R.; Skoblo, R.; Russell, R.B.; Bourne, P.E.; Bork, P.; Preissner, R. SuperTarget and Matador: resources for exploring drug-target relationships. *Nucleic Acids Res.*, **2008**, *36*, D919-D922.

[11] Hecker, N.; Ahmed, J.; Von Eichborn, J.; Dunkel, M.; Macha, K.; Eckert, A.; Gilson, M.K.; Bourne, P.E.; Preissner, R. SuperTarget goes quantitative: update on drug-target interactions. *Nucleic Acids Res.*, **2012**, *40*, D1113-1117.

[12] Li, Y.H.; Yu, C.Y.; Li, X.X.; Zhang, P.; Tang, J.; Yang, Q.; Fu, T.; Zhang, X.; Cui, X.; Tu, G.; Zhang, Y.; Li, S.; Yang, F.; Sun, Q.; Qin, C.; Zeng, X.; Chen, Z.; Chen, Y.Z.; Zhu, F. Therapeutic target database update 2018: Enriched resource for facilitating bench-to-clinic research of targeted therapeutics. *Nucleic Acids Res.,* **2018**, *46*(D1), D1121-D1127.

[13] Gilson, M.K.; Liu, T.; Baitaluk, M.; Nicola, G.; Hwang, L.; Chong, J. BindingDB in 2015: A public database for medicinal chemistry,

computational chemistry and systems pharmacology. *Nucleic Acids Res.*, **2016**, *44*, D1045-D1053.

[14] Placzek, S.; Schomburg, I.; Chang, A.; Jeske, L.; Ulbrich, M.; Tillack, J.; Schomburg, D. Brenda in 2017: New perspectives and new tools in BRENDA. *Nucleic Acids Res.*, **2017**, *45*, D380-D388.

[15] Kim, S.; Thiessen, P.A.; Bolton, E.E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B.A. PubChem Substance and Compound databases. *Nucleic Acids Res.,* **2016**, *44*, 1202-1213.

[16] Consortium, T.U. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.,* **2017**, *45*, D158-D169.

[17] Chatr-Aryamontri, A.; Oughtred, R.; Boucher, L.; Rust, J.; Chang, C.; Kolas, N.K.; O'Donnell, L.; Oster, S.; Theesfeld, C.; Sellam, A.; Stark, C.; Breitkreutz, B.J.; Dolinski, K.; Tyers, M. The BioGRID interaction database: 2017 update. *Nucleic Acids Res.,* **2017**, *45*, D369-D379.

[18] Kuhn, M.; Letunic, I.; Jensen, L.J.; Bork, P. The SIDER database of drugs and side effects. *Nucleic Acids Res.*, **2016**, *44*, D1075-1079.

[19] Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T. Gene Ontology: Tool for the unification of biology. *Nat. Genet.*, **2000**, *25*, 25-29.

[20] Finn, R.D.; Coggill, P.; Eberhardt, R.Y.; Eddy, S.R.; Mistry, J.; Mitchell, A.L.; Potter, S.C.; Punta, M.; Qureshi, M.; Sangrador-Vegas, A. The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Res.* **2016**, *44*, D279-D285.

[21] Willighagen, E.L.; Mayfield, J.W.; Alvarsson, J.; Berg, A.; Carlsson, L.; Jeliazkova, N.; Kuhn, S.; Pluskal, T.; Rojas-Cherto, M.; Spjuth, O.; Torrance, G.; Evelo, C.T.; Guha, R.; Steinbeck, C. Erratum to: The Chemistry Development Kit (CDK) v2.0: Atom typing, depiction, molecular formulas, and substructure searching. *J. Cheminform.,* **2017**, *9*, 53.

[22] Yap, C.W. PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. *J. Comput. Chem.,* **2011**, *32*, 1466-1474.

[23] O'Boyle, N.M.; Banck, M.; James, C.A.; Morley, C.; Vandermeersch, T.; Hutchison, G.R. Open Babel: An open chemical toolbox. *J. Cheminform.,* **2011**, *3*, 33.

[24] Hinselmann, G.; Rosenbaum, L.; Jahn, A.; Fechner, N.; Zell, A. jCompoundMapper: An open source Java library and command-line tool for chemical fingerprints. *J. Cheminform.,* **2011**, *3*, 3.

[25] Cao, D.S.; Xu, Q.S.; Hu, Q.N.; Liang, Y.Z. ChemoPy: freely available python package for computational biology and chemoinformatics. *Bioinformatics*, **2013**, *29*, 1092-1094.

[26] Dong, J.; Cao, D.S.; Miao, H.Y.; Liu, S.; Deng, B.C.; Yun, Y.H.; Wang, N.N.; Lu, A.P.; Zeng, W.B.; Chen, A.F. ChemDes: an integrated web-based platform for molecular descriptor and fingerprint computation. *J. Cheminform.,* **2015**, *7*, 60.

[27] Liu, B.; Liu, F.; Fang, L.; Wang, X.; Chou, K.-C. repDNA: A Python package to generate various modes of feature vectors for DNA sequences by incorporating user-defined physicochemical properties and sequence-order effects. *Bioinformatics*, **2014**, *31*, 1307-1309.

[28] Liu, B.; Liu, F.; Fang, L.; Wang, X.; Chou, K.-C. repRNA: A web server for generating various feature vectors of RNA sequences. *Mol. Genet. Genomics*, **2016**, *291*, 473-481.

[29] Liu, B. BioSeq-Analysis: A platform for DNA, RNA and protein sequence analysis based on machine learning approaches. *Brief. Bioinform.*, **2017**, doi: 10.1093/bib/bbx165.

[30] Chen, W.; Feng, P.-M.; Lin, H.; Chou, K.-C. iSS-PseDNC: Identifying splicing sites using pseudo dinucleotide composition. *BioMed Res. Int.*, **2014**, *2014*, 623149.

[31] Chen, W.; Yang, H.; Feng, P.; Ding, H.; Lin, H. iDNA4mC: Identifying DNA N4-methylcytosine sites based on nucleotide chemical properties. *Bioinformatics*, **2017**, *33*, 3518-3523.

[32] Liu, X.; Ding, J.; Gong, F. piRNA identification based on motif discovery. *Mol. Biosyst.*, **2014**, *10*, 3075-3080.

[33] Li, D.; Luo, L.; Zhang, W.; Liu, F.; Luo, F. A genetic algorithm-based weighted ensemble method for predicting transposon-derived piRNAs. *BMC Bioinformatics*, **2016**, *17*, 329.

[34] Luo, L.; Li, D.; Zhang, W.; Tu, S.; Zhu, X.; Tian, G. Accurate prediction of transposon-derived piRNAs by integrating various sequential and physicochemical features. *Plos One*, **2016**, *11*, e0153268.

[35] Rao, H.; Zhu, F.; Yang, G.; Li, Z.; Chen, Y.Z. Update of PROFEAT: A web server for computing structural and physicochemical features of proteins and peptides from amino acid sequence. *Nucleic Acids Res.,* **2006**, *39*, 385-390.

[36] Chen, W.; Zhang, X.; Brooker, J.; Lin, H.; Zhang, L.; Chou, K.-C. PseKNC-General: A cross-platform package for generating various modes of pseudo nucleotide compositions. *Bioinformatics*, **2014**, *31*, 119-120.

[37] Liu, B.; Liu, F.; Wang, X.; Chen, J.; Fang, L.; Chou, K. Pse-in-One: A web server for generating various modes of pseudo components of DNA, RNA, and protein sequences. *Nucleic Acids Res.,* **2015**, *43*, W65-W71.

[38] Zhang, W.; Qu, Q.; Zhang, Y.; Wang, W. The linear neighborhood propagation method for predicting long non-coding RNA-protein interactions. *Neurocomputing*, **2018**, *273*, 526-534.

[39] Zhang, W.; Yue, X.; Liu, F.; Chen, Y.; Tu, S.; Zhang, X. A unified frame of predicting side effects of drugs by using linear neighborhood similarity. *BMC Syst. Biol.*, **2017**, *11*, 101.

[40] Zhang, W.; Chen, Y.; Tu, S.; Liu, F.; Qu, Q. Drug side effect prediction through linear neighborhoods and multiple data source integration. *IEEE Int. Conf. Bioinformatics Biomed. (BIBM)*, **2016**, 427-434.

[41] Zhang, W.; Chen, Y.; Liu, F.; Luo, F.; Tian, G.; Li, X. Predicting potential drug-drug interactions by integrating chemical, biological, phenotypic and network data. *BMC Bioinformatics*, **2017**, *18*, 18.

[42] Hattori, M.; Tanaka, N.; Kanehisa, M.; Goto, S. SIMCOMP/SUBCOMP: Chemical structure search servers for network analyses. *Nucleic Acids Res.*, **2010**, *38*, W652-W656.

[43] Zou, Q.; Hu, Q.; Guo, M.; Wang, G. HAlign: Fast multiple similar DNA/RNA sequence alignment based on the centre star strategy. *Bioinformatics*, **2015**, *31*, 2475-2481.

[44] Su, W.; Liao, X.; Lu, Y.; Zou, Q.; Peng, S. Multiple sequence alignment based on a suffix tree and center-star strategy: A linear method for multiple nucleotide sequence alignment on spark parallel framework. *J. Comput. Biol.*, **2017**, *24*, 1230-1242.

[45] Wan, S.; Zou, Q. HAlign-II: Efficient ultra-large multiple sequence alignment and phylogenetic tree reconstruction with distributed and parallel computing. *Algorithms Mol. Biol.*, **2017**, *12*, 25.

[46] Bleakley, K.; Yamanishi, Y. Supervised prediction of drug-target interactions using bipartite local models. *Bioinformatics*, **2009**, *25*, 2397-2403.

[47] Yamanishi, Y.; Kotera, M.; Kanehisa, M.; Goto, S. Drug-target interaction prediction from chemical, genomic and pharmacological data in an integrated framework. *Bioinformatics*, **2010**, *26*, i246-i254.

[48] Tabei, Y.; Pauwels, E.; Stoven, V.; Takemoto, K.; Yamanishi, Y. Identification of chemogenomic features from drug-target interaction networks using interpretable classifiers. *Bioinformatics*, **2012**, *28*, i487-i494.

[49] Mei, J.P.; Kwoh, C.K.; Yang, P.; Li, X.L.; Zheng, J. Drug-target interaction prediction by learning from local information and neighbors. *Bioinformatics*, **2013**, *29*, 238-245.

[50] Liu, H.; Sun, J.; Guan, J.; Zheng, J.; Zhou, S. Improving compound-protein interaction prediction by building up highly credible negative samples. *Bioinformatics*, **2015**, *31*, i221-i229.

[51] Mousavian, Z.; Khakabimamaghani, S.; Kavousi, K.; Masoudi-Nejad, A. Drug-target interaction prediction from PSSM based evolutionary information. *J. Pharmacol. Toxicol. Methods*, **2015**, *78*, 42-51.

[52] Ezzat, A.; Wu, M.; Li, X.L.; Kwoh, C.K. Drug-target interaction prediction *via* class imbalance-aware ensemble learning. *BMC Bioinformatics*, **2016**, *17*, 509.

[53] Ding, Y.J.; Tang, J.J.; Guo, F. Identification of drug-target interactions *via* multiple information integration. *Inf. Sci.*, **2017**, *418*, 546-560.

[54] Keum, J.; Nam, H. SELF-BLM: Prediction of drug-target interactions *via* self-training SVM. *Plos One*, **2017**, *12*, 16.

[55] Li, Z.; Han, P.; You, Z.H.; Li, X.; Zhang, Y.; Yu, H.; Nie, R.; Chen, X. *In silico* prediction of drug-target interaction networks based on drug chemical structure and protein sequences. *Sci. Rep.,* **2017**, *7*, 11174.

[56] Meng, F.R.; You, Z.H.; Chen, X.; Zhou, Y.; An, J.Y. Prediction of drug-target interaction networks from the integration of protein sequences and drug chemical structures. *Molecules*, **2017**, *22*, E1119.

[57] Peng, L.H.; Zhu, W.; Liao, B.; Duan, Y.; Chen, M.; Chen, Y.; Yang, J.L. Screening drug-target interactions with positive-unlabeled learning. *Sci. Rep.,* **2017**, *7*, 17.

[58] Chen, X.; Liu, M.X.; Yan, G.Y. Drug-target interaction prediction by random walk on the heterogeneous network. *Mol. Biosyst.*, **2012**, *8*, 1970-1978.

[59] Cheng, F.; Liu, C.; Jiang, J.; Lu, W.; Li, W.; Liu, G.; Zhou, W.; Huang, J.; Tang, Y. Prediction of drug-target interactions and drug repositioning via network-based inference. *PLoS Comput. Biol.*, **2012**, *8*, e1002503.

[60] Alaimo, S.; Pulvirenti, A.; Giugno, R.; Ferro, A. Drug-target interaction prediction through domain-tuned network-based inference. *Bioinformatics*, **2013**, *29*, 2004-2008.

[61] Emig, D.; Ivliev, A.; Pustovalova, O.; Lancashire, L.; Bureeva, S.; Nikolsky, Y.; Bessarabova, M. Drug target prediction and repositioning using an integrated network-based approach. *PLoS One*, **2013**, *8*, e60618.

[62] Re, M.; Valentini, G. Network-based drug ranking and repositioning with respect to DrugBank therapeutic categories. *IEEE/ACM Trans. Comput. Biol. Bioinform.*, **2013**, *10*, 1359-1371.

[63] Yu, W.; Yan, Y.; Liu, Q.; Wang, J.; Jiang, Z. Predicting drug-target interaction networks of human diseases based on multiple feature information. *Pharmacogenomics*, **2013**, *14*, 1701-1707.

[64] Alaimo, S.; Bonnici, V.; Cancemi, D.; Ferro, A.; Giugno, R.; Pulvirenti, A. DT-Web: A web-based application for drug-target interaction and drug combination prediction through domain-tuned network-based inference. *BMC Syst. Biol.*, **2015**, *9*(Suppl 3), S4.

[65] Seal, A.; Ahn, Y.Y.; Wild, D.J. Optimizing drug-target interaction prediction based on random walk on heterogeneous networks. *J. Cheminform.,* **2015**, *7*, 40.

[66] Yan, X.Y.; Zhang, S.W.; Zhang, S.Y. Prediction of drug-target interaction by label propagation with mutual interaction information derived from heterogeneous network. *Mol. Biosyst.*, **2016**, *12*(2), 520-531.

[67] Fu, G.; Ding, Y.; Seal, A.; Chen, B.; Sun, Y.Z.; Bolton, E. Predicting drug target interactions using meta-path-based semantic network analysis. *BMC Bioinformatics*, **2016**, *17*, 10.

[68] Li, Z.C.; Huang, M.H.; Zhong, W.Q.; Liu, Z.Q.; Xie, Y.; Dai, Z.; Zou, X.Y. Identification of drug-target interaction from interactome network with 'guilt-by-association' principle and topology features. *Bioinformatics*, **2016**, *32*, 1057-1064.

[69] Lu, Y.D.; Guo, Y.F.; Korhonen, A. Link prediction in drug-target interactions network using similarity indices. *BMC Bioinformatics*, **2017**, *18*, 9.

[70] Luo, Y.; Zhao, X.; Zhou, J.; Yang, J.; Zhang, Y.; Kuang, W.; Peng, J.; Chen, L.; Zeng, J. A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. *Nat. Commun.,* **2017**, *8*, 573.

[71] Wu, Z.R.; Cheng, F.X.; Li, J.; Li, W.H.; Liu, G.X.; Tang, Y. SDTNBI: An integrated network and chemoinformatics tool for systematic prediction of drug-target interactions and drug repositioning. *Brief. Bioinform.*, **2017**, *18*, 333-347.

[72] Gonen, M. Predicting drug-target interactions from chemical and genomic kernels using Bayesian matrix factorization. *Bioinformatics*, **2012**, *28*, 2304-2310.

[73] Zheng, X.; Ding, H.; Mamitsuka, H.; Zhu, S. In: *Collaborative Matrix Factorization With Multiple Similarities For Predicting Drug-Target Interactions*, Proceedings of the 19th ACM SIGKDD international conference on knowledge discovery and data mining, **2013**.

[74] Liu, Y.; Wu, M.; Miao, C.; Zhao, P.; Li, X.L. Neighborhood regularized logistic matrix factorization for drug-target interaction prediction. *PLoS Comput. Biol.,* **2016**, *12*, e1004760.

[75] Hao, M.; Bryant, S.H.; Wang, Y.L. Predicting drug-target interactionsby dual-network integrated logistic matrix factorization. *Sci. Rep.*, **2017**, *7*, 11.

[76] Ezzat, A.; Zhao, P.L.; Wu, M.; Li, X.L.; Kwoh, C.K. Drug-target interaction prediction with graph regularized matrix factorization. *IEEE-ACM Trans. Comput. Biol. Bioinform.*, **2017**, *14*, 646-656.

[77] Peska, L.; Buza, K.; Koller, J. Drug-target interaction prediction: A Bayesian ranking approach. *Comput. Methods Programs Biomed.,* **2017**, *152*, 15-21.

[78] Hu, P.W.; Chan, K.C.C.; You, Z.H. Large-scale prediction of drug-target interactions from deep representations. In: *2016 International Joint Conference on Neural Networks*, IEEE: New York, 2016; pp. 1236-1243.

[79] Yuan, Q.; Gao, J.; Wu, D.; Zhang, S.; Mamitsuka, H.; Zhu, S. DrugE-Rank: Improving drug-target interaction prediction of new candidate drugs or targets by ensemble learning to rank. *Bioinformatics*, **2016**, *32*, i18-i27.

[80] Tian, K.; Shao, M.; Wang, Y.; Guan, J.; Zhou, S. Boosting compound-protein interaction prediction by deep learning. *Methods*, **2016**, *110*, 64-72.

[81] Wen, M.; Zhang, Z.; Niu, S.; Sha, H.; Yang, R.; Yun, Y.; Lu, H. Deep-learning-based drug-target interaction prediction. *J. Proteome Res.*, **2017**, *16*, 1401-1409.

[82] Zong, N.S.; Kim, H.; Ngo, V.; Harismendy, O. Deep mining heterogeneous networks of biomedical linked data to predict novel drug-target associations. *Bioinformatics*, **2017**, *33*, 2337-2344.

[83] Jacob, L.; Vert, J.P. Protein-ligand interaction prediction: An improved chemogenomics approach. *Bioinformatics*, **2008**, *24*, 2149-2156.

[84] Xia, Z.; Wu, L.Y.; Zhou, X.; Wong, S.T. Semi-supervised drug-protein interaction prediction from heterogeneous biological spaces. *BMC Syst. Biol.,* **2010**, *4*(Suppl 2), S6.

[85] van Laarhoven, T.; Nabuurs, S.B.; Marchiori, E. Gaussian interaction profile kernels for predicting drug-target interaction. *Bioinformatics*, **2011**, *27*, 3036-3043.

[86] Shang, F.; Jiao, L.C.; Liu, Y. Integrating spectral kernel learning and constraints in semi-supervised classification. *Neural Process. Lett.*, **2012**, *36*, 101-115.

[87] Nascimento, A.C.; Prudencio, R.B.; Costa, I.G. A multiple kernel learning algorithm for drug-target interaction prediction. *BMC Bioinformatics*, **2016**, *17*, 46.

[88] Wang, Y.; Zeng, J. Predicting drug-target interactions using restricted Boltzmann machines. *Bioinformatics*, **2013**, *29*, i126-i134.

[89] Koohi, A. Prediction of drug-target interactions using popular collaborative filtering methods. In: *2013 Ieee International Workshop on Genomic Signal Processing and Statistics*, IEEE: New York, 2013; pp. 58-61.

[90] Fakhraei, S.; Huang, B.; Raschid, L.; Getoor, L. Network-based drug-target interaction prediction with probabilistic soft logic. *IEEE/ACM Trans Comput. Biol. Bioinform.,* **2014**, *11*, 775-787.

[91] Zhang, X.; Li, L.; Ng, M.K.; Zhang, S. Drug-target interaction prediction by integrating multiview network data. *Comput. Biol. Chem.,* **2017**, *69*, 185-193.

[92] Zhang, W.; Chen, Y.; Li, D. Drug-target interaction prediction through label propagation with linear neighborhood information. *Molecules*, **2017**, *22*, 2056.