International Journal of Advanced Scientific Research
www.allscientificjournal.com
ISSN: 2456-0421
Received: 10-05-2021; Accepted: 25-05-2021; Published: 09-06-2021
Volume 6, Issue 3, 2021, Page No. 21-25

Artificial intelligence in drug-target interaction prediction for its potential applications in personalized medicine

Meghna Goswami[1], Priyanka Patel[2*], Shubhalaxmi Sahoo[3]
[1] Research Assistant, Rapture Biotech, Ahmedabad, Gujarat, India
[2] Director, Rapture Biotech, Ahmedabad, Gujarat, India
[3] Research Intern, Rapture Biotech, Ahmedabad, Gujarat, India

Abstract
The post human genome project era has opened new avenue in the field of drug discovery which depends greatly over the massive data generated from the genomic and transcriptomic sequencing projects, across the globe, specifically based on various genetic and metabolic disorders. With the developments in the computational performances and algorithmic efficiency, now it is becoming possible to analyze the huge sequencing data. This review targets the machine learning and network based techniques of Artificial Intelligence (AI) for the prediction of drug target interaction (DTI) which is utmost essentiality for drug discovery aiming personalized medicine or individualized therapeutics. Personalized medicine is the new generation medicine practice with an obligation to develop systemic drug's efficacy and dosage as per individual requirements. This personalized therapeutic dosage and effects of drug molecules can be achievable by the study of DTI research with the use of AI.

Keywords: drug discovery, artificial intelligence, personalized medicine, drug target interaction

Introduction

Personalized medicine is the new era of medicine. It is greatly facilitated by the data generated by the research studies carried from all around the world in the area of human genome sequencing and insights developed from the understanding of this data. These advancements in the understanding of the human genome are the creating the basics of personalized medicine and would transform the working of healthcare sector based on systemic drug requirements of individuals. In personalized medicine development, use of marker-assisted diagnosis and targeted therapies derived from an individual molecular profile, and the way drugs react within the human body are developed and therapeutics are managed accordingly.

Drug discovery and development has a traditionally been a linear process. The adoption of a personalized medicine strategy in drug discovery and development necessitates a paradigm shift from a linear process to an integrated and heuristic one. Drug discovery is the most significant task in the development of personalized medicine. In drug discovery, drug–target interaction (DTI) always plays an important role. However, prediction of a DTI experimentally is a time-consuming and costly affair. To overcome this barrier and make the research and study of DTI efficient, various *in silico* methods have been developed over several decades. Recently, as an advancement *in silico* approaches for analyzing, understanding the complexity of processes in biological systems, parallel computing and artificial intelligence (AI) are in use. These *in silico* methods are also being utilized for the prediction of DTI [1].

On the contrary, development of system biology and network pharmacology, has changed the drug discovery pattern from the linear mode of "one-drug one-target one-disease" to network mode of "multi-drugs multi-targets multi-diseases" [2]. In the network mode, the single drug acts on multiple targets *in vivo*, instead of binding to a single target and the outcomes are complied accordingly [3]. The poly-pharmacological profiles for a drug lead to both required therapeutic effects along with other safety problems and side effects can also be minimized based on the outcomes of these single drug-multi target models. Therefore, identifying interactions between drug and target is an basic essentiality of modern drug discovery, which aids the therapeutic efficacy along with minimal non target interactions [4].

Advancement in computational methods has made it possible to deliver crucial methods for prediction of DTIs making it resource efficient. For identification of DTIs experimentally we determine the inhibition constant (Ki), dissociation constant (Kd), half-maximal inhibitory concentration (IC50) or half-maximal effective concentration (EC50) values between drugs and target proteins by *in-vitro* or *in-vivo* assays. It is tedious and expensive to determine all possible DTIs experimentally and systematically *in vivo* simultaneously due to resource limitations [6]. Inversely the *in-vitro* set-ups use already generated data and *in silico* approaches for the inference of useful information with countless variations to provide the results. The results of prediction divide the methods into two categories, qualitative (i.e., classification) and quantitative (i.e., regression) methods [7]. These methods are divided into various classes, which include molecular docking-based, pharmacophore-based, similarity-based, machine learning-based and network-based methods, even though their concepts could have common characteristics between each other [8].

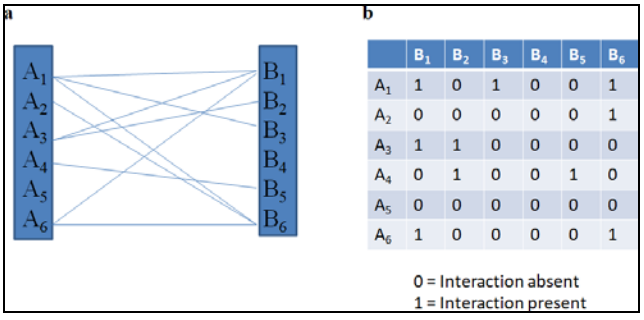
Role of Artificial Intelligence in Drug discovery

An ideal drug discovery for a particular disease that can have many treatment strategies and can consider the entire relevant medical information from a patient, including genomics and/ or proteomics data integrated with features describing drug's properties. It is important to predict drug's interactions with biomolecules such as proteins, DNA, RNA, miRNA and others for the better and accurate outcomes [9]. A combination of these approaches is necessary for clinical applications of AI based predictive models, especially for complex diseases. Drug-target interactions prediction is a smaller part of the monotherapy optimization problem, but it is no less important. Prediction of DTIs is vital to the repurposing of already existing drugs and search for novel drugs. Study of DTI can also be utilized to develop an understanding of complicated signaling and metabolic pathways involved in the drug metabolism. The prediction of DTI presented here relies mainly on two branches of AI techniques:

1. Network theory approaches, in which the input characterizes the protein networks with which a drug interacts [10] and
2. Machine learning approaches, which utilizes architectures like Support Vector Machine (SVM) [11].

Network Theory Approach for drug target interaction

Network-based inference (NBI) is used for predicting potential of DTIs with help of existing DTIs network (i.e., positive samples), without any information regarding chemical structures, protein structures or sequences. In recent times, various network-based methods have already been developed [12]. The advantages of these methods are independent from tertiary structures of targets, and this enables them to cover more target space. Network based methods are quick and easy. Potential DTIs can be predicted by performing simple processes like resource diffusion [13], mutual filtering [14], and random walk on networks. One can represent network by using matrices by doing matrix multiplication mathematically [15, 16]. NBI uses recommendation algorithms for deriving these methods [17] and link predicting the algorithms in composite networks [19]. In DTI network based inference a heterogeneous network is built having a collection of two sets of data which interacts with each other [20]. The interactions between sets are linked by connecting nodes of the network and are represented as a matrix and are denoted as interaction, adjacency, or connectivity matrix [21]. It builds models utilizing a number of similarity-based modeling mechanisms such as drug-based similarity inference, target-based similarity inference and a bipartite graph network-based approach, also termed as network-based inference (NBI), for predicting various drug-target interactions. NBI has been proved to be the most powerful method and was used to test further data. [22-24]



	B_1	B_2	B_3	B_4	B_5	B_6
A_1	1	0	1	0	0	1
A_2	0	0	0	0	0	1
A_3	1	1	0	0	0	0
A_4	0	1	0	0	1	0
A_5	0	0	0	0	0	0
A_6	1	0	0	0	0	1

Fig 1: Representation of NBI method by bipartite graph showing interactions

A bipartite graph is a representation of method in which vertices, illustrating interactions, are drawn only between disjoint sets (Figure 1). Interaction between objects in disjoint sets A_1 and B_1 represents connection by blue line (Figure 1-a). The disjoint sets are represented as protein which interacts with the other protein of different set but not within the set. A protein in set B will not interact with another protein in the set B but may interact with any or all proteins in set A. A bipartite graph is generated from adjacent matrix (Figure 1-b), where 1 represents the interaction between the two elements. In Figure 1, A_5 and B_4 appear in the matrix but do not show interaction as value in the matrix is 0 in their representative row and column, so it does not show any interaction in bipartite graph [1].

On the basis of these interaction predictions, one can follow two learning strategies: Local approach and Global approach. Local approach for predicting interaction which handles two set of the DTI network separately. It first divides DTI network into different feature sets, the drug-based set and the protein-based set. For each set of learning, tasks are tackled separately and result output is unified. In the absence of information on one or both sides, generally local models are built on a single feature space, ligand (drug) space or target (protein) space [25]. Ligand-based models are built on the known ligands that interact with targets (proteins). However, performance of these models is impaired when it comes to target proteins with only a small number (or even none) of known binding ligands are to be studied [20]. Alternatively, target-based models are built on the target proteins using protein (3D) structure information. Nevertheless, the 3D structure of many target proteins is often unavailable [26, 27]. There are various method of network based theories, which helps in prediction of the interactions between drug and target molecules as mentioned further.

a. Network-based inference (NBI),
b. Substructure-drug-target network-based inference (SDTNBI),
c. Balanced substructure-drug-target network-based inference (bSDTNBI) [13].

Network-based inferences do not depend on tertiary structure of targets or negative samples while comparing complex structure-based and machine learning-based methods. The *in-silico* method for prediction of interaction between drug and target are effortless and easy. Also, network-based methods are rapid as it is performed *in-silico* [28].

Machine learning using Support Vector Machine (SVM) for classification of similar drugs

Machine learning comprises of two models supervised learning model and unsupervised learning model. When the data being used are labeled, supervised learning is used and when the data are not labeled unsupervised learning is used. Both of these supervised and unsupervised learning models, cluster the data into groups followed by mapping new data in to the groups formed. Support vector machine are supervised learning models associated with learning algorithms that analyze data used for classification and regression analysis. Machine learning using the SVM is simple to use when compared to Neural Network. Machine learning utilizing SVM, mainly aims to separate categorical variables by drawing multidimensional surface between them as depicted in Figure 3. In drug-target interaction,

SVM is used for classifying drugs that have therapeutically similar side effects but are not structurally similar [29]. For example, Cetrizine and Acitretin, in spite of their structural differences both inhibit the same target (i.e., HRH1) and depict the same therapeutic ability (Figure 2).

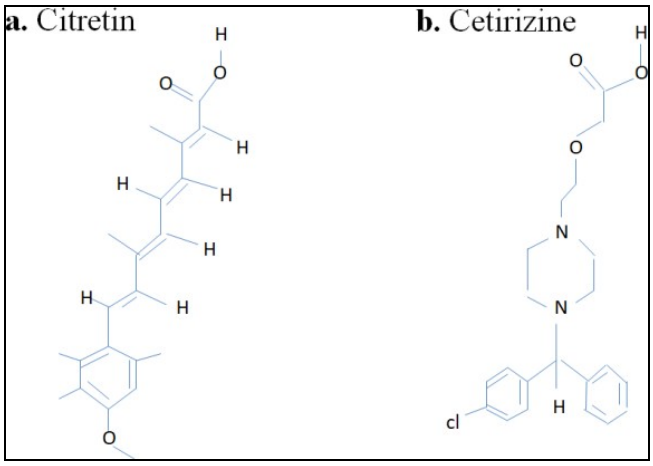


Fig 2: Structure of drugs (a) Acitretin and (b) Cetirizine

Support vector machine will separate data sets in two categories by mapping point in space. This can be achieved by SVM training algorithm which builds model that will assign a data into two categories, making it a non-probabilistic binary linear classifier. Training model will separate the trained data set in separate categories which are divided by a clear gap that will be as wide as possible. The new test data sets are then mapped into that same space and predictions of its category are based on which side of the gap they fall [30]. SVM has efficiency to perform both linear classification and nonlinear classification using kernel trick, by implicitly mapping inputs into high dimensional feature space. A representative example data set is being classified into two classes (Triangles and circles) by generating boundaries (Figure 3). The category boundaries are represented by solid lines and are separated by the maximum margin linear classifier (dashed line). The accuracy of model can be determined by the number of molecules placed on the correct side of the line.

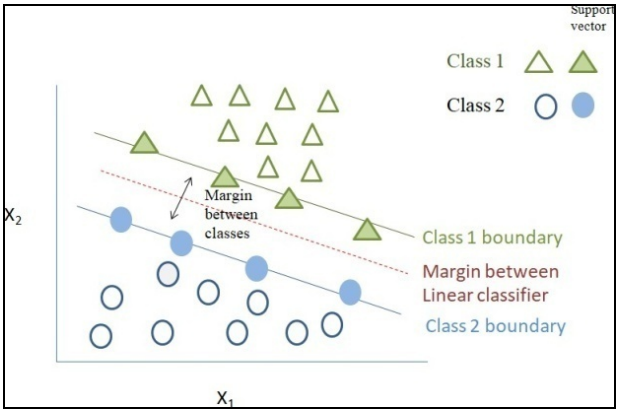


Fig 3: Classification of data set into classes by constructing margin between two classes using linear classifier

AI and Personalized medicine

System biology and network pharmacology has shown development and proposed various computational methods for DTIs prediction with high accuracy and low cost thus increasing the efficiency. Artificial intelligence has shown greater advantages in DTI based research for the predictions of the nearly accurate outcomes as compared to the wet lab experiments and made it accessible to analyze and infer the results for the clinical application. Based on these virtues, AI can be an inseparable part of the study of DTIs, which are considered the base of the personalized medicine. Figure 3 has been designed to describe the interconnections between personalized medicine and AI technologies based on all the phases of the development (T0-T4) of drug molecules which can be used for personalized medicine.

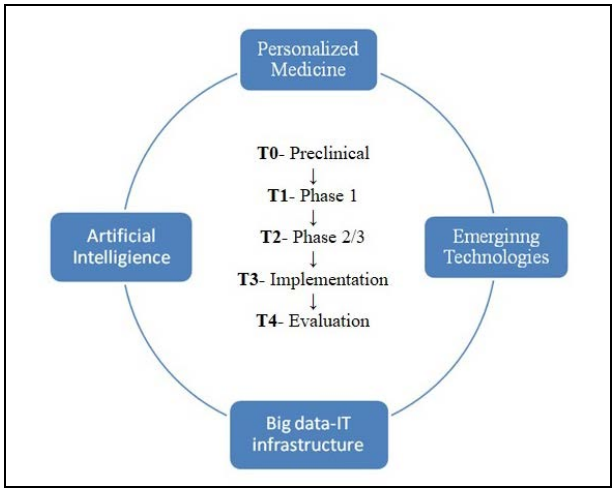


Fig 3: Four emerging themes in biomedical science

The development of personalized medicine, as with the diagnosis and the treatment of patients, proceeds in different phase i.e., from T1 to T4. There are different ways of defining and referring to these phases, however all of them point to opportunity for AI to have a significant impact on personalized medicine. In general drug development for a population the pattern is as described from T0 to T4, whereas for personalized medicine, it would needs to be designed as P0 to P4. (Table 1) This phase can command AI techniques if the right data and stimulus is present. [31]

Table 1: An analogous representation of the stages in development of drug and the stages to be covered

General Drug Development Population Benefits	Personalized Medicine Patient benefits
T0 (Discovery/preclinical) Target identification- Validation	P0 (Diagnosis) Genomic/ Biochemical/ Physical Assesment of interactions
T1 (First-in-Human/Proof-of-concept) Dose Ranging- Safety Studies Patient Engagement	P1 (Intervention Point Assessment) Molecular Physiological Studies
T2 (Clinical Assesment) Population Studies and Trials Monitoring outcomes	P2 (Intervention Choice) Drug/ Intervention Repurposing
T3 (Clinical Implementation) Development Studies/ Optimization and work on Surveys	P3 (Intervention Testing) Design of N-of-1 Clinical test Drug monitoring
T4 (Post-Development Evaluation) Acceptance and Experience, Refinement and adaption of regulatory issues	P4 (Data Warehousing the results) Merge into Queryable Database, Open source

Personalized medicine revolutionizing drug discovery

With the beginning of post-genome era, there is medicine practices which are being developed based on ethnicities. There are differential guidelines for the practice of medicine based on the genetic origin of the population as defined blood glucose levels, Body mass index, height to weight ratio etc. This has also opened the doors to the newly emerging branch of personalized medicine, which is being supported greatly by the availability of the large sequencing data sets of human and its biotic environment. There have been many recent developments of *in silico* techniques which has made possible to mine the useful data and gain insightful results which can be applied clinically to population masses. Moving a step ahead, if these sequencing techniques are employed for the individuals, there would be revolutionary changes in the world of medicine, which might lead to the growth of personalized medicine encompassing molecular diagnostics, pharmacogenomics, targeted clinical therapies and use of risk algorithms to improve health care. General drug discovery has been a linear process (Figure 5) since ages. The advancements in AI based DTIs have opened new pavements for personalized medicine strategies. These newly surfaced technologies of AI based DTIs are shifting the new drug discovery and research from linear process to an integrated and a complex one (Figure 6). This amelioration of AI in drug discovery and designing for the development of personalized medicine will lead to a series of new research pathways. The beginning of the drug discovery can be termed as the selection and validation of drug and its target, screening of molecules with the help of AI, and *in silico* based research. The drug molecules with positive results and fewer side effects in AI approaches can be chosen to complete the phases of preclinical assessment resulting in the development of validated clinical drug molecule [32].

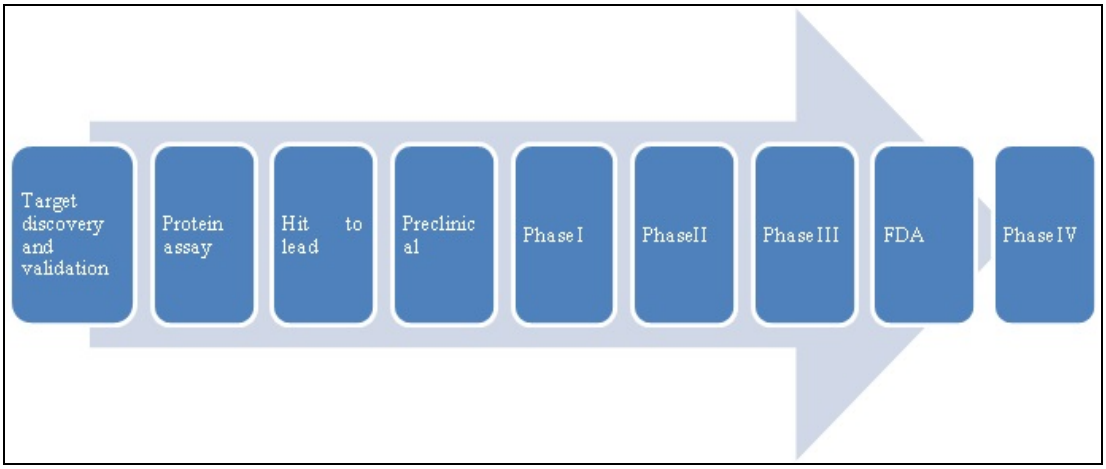


Fig 5: General drug discovery a linear process

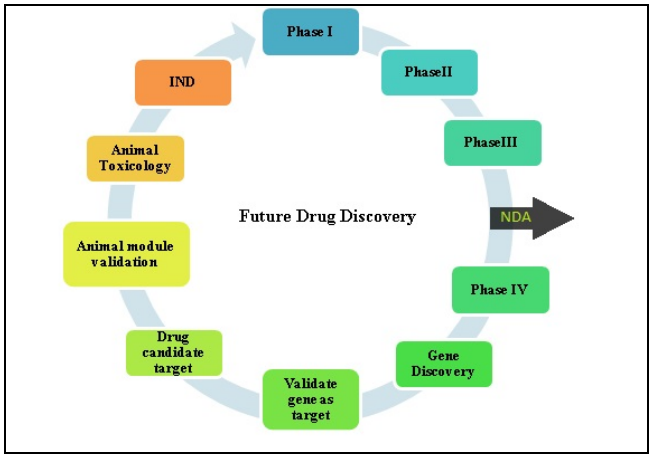


Fig 6: Future drug discovery – an integrated process (IND- investigational new drug; NDA- new drug application)

Future Aspects

Improvement of DTIs based prediction of the outcomes is possible when appropriate AI models are used. New prediction algorithms including hierarchical structure stochastic block and likelihood analysis can be developed for gaining a better related outcome. Researchers can also add multi-scale biomedical data, which includes drug side-effect associations; drug-induced gene expression profiles protein-protein interactions, ADMET properties, clinical data etc. Machine learning and network-based approaches have been utilized in many fields adequately and efficiently. There is a need for additional joint efforts from experts in fields of computational biology, informational technology, molecular biology and clinical practice on a single front, to carry out multidisciplinary research and infer better results from innumerous scratch data for the development of technologies in personalized medicine.

Conclusion

Artificial intelligence undoubtedly has numerous benefits in DTI prediction as compared to the work being carried out *in vivo*. DTIs prediction *in vivo* is certainly not resource efficient. Development of system biology and network pharmacology has changed the drug discovery and made the work easier. AI has a big role to play in the future of this ever expanding digital world. But, it has certain limitations. AI methods are non-quantitative and only provide score for possible DTIs, where high score indicates highest probability of drug interaction with its target. The binding affinity of the predicted drug target is unknown, hence, research studies finding the binding affinity of the drug and target molecules is also required. AI methods can be used for DTIs prediction for approved drug and chemical entities; with additional efforts in research of DTIs prediction for the targets whose ligands are not present. AI with DTI prediction provides an alternative tool for studies in drug repurposing, new drug discovery, system pharmacology and toxicology.

References

1. Romm E and Tsigelny I. Artificial Intelligence in Drug Treatment. Annual Review of Pharmacology and Toxicology,2020:60(1):353-369.
2. Medina-Franco J, Giulianotti M, Welmaker G, Houghten R. Shifting from the single to the multitarget paradigm in drug discovery. Drug Discovery Today,2013:18(9-10):495-501.
3. Hopkins A. Network pharmacology: the next paradigm in drug discovery. Nature Chemical Biology,2008:4(11):682-690.
4. Anighoro A, Bajorath J, Rastelli G. Polypharmacology: Challenges and Opportunities in Drug Discovery. Journal of Medicinal Chemistry,2014:57(19):7874-7887.
5. Zheng M, Liu X, Xu Y, Li H, Luo C, Jiang H. Computational methods for drug design and discovery: focus on China. Trends in Pharmacological Sciences,2013:34(10):549-559.
6. Chen X, Yan C, Zhang X, Zhang X, Dai F, Yin J *et al*. Drug–target interaction prediction: databases, web servers and computational models. Briefings in Bioinformatics,2015:17(4):696-712.
7. Lavecchia A, Cerchia C. In silico methods to address polypharmacology: current status, applications and future perspectives. Drug Discovery Today,2016:21(2):288-298.
8. 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 Computational Biology,2012:8(5):e1002503.
9. Smith T and Waterman M. Identification of common molecular subsequences. Journal of Molecular Biology,1981:147(1):195-197.
10. Wu Z, Lu W, Wu D, Luo A, Bian H, Li J *et al*. In silico prediction of chemical mechanism of action via an improved network-based inference method. British Journal of Pharmacology,2016:173(23):3372-3385.
11. Cortes C and Vapnik V. Support-vector networks. Machine learning,1995:20(3):273-97.
12. Xiao X, Min J, Lin W, Liu Z, Cheng X, Chou K. iDrug-Target: predicting the interactions between drug compounds and target proteins in cellular networking via benchmark dataset optimization approach. Journal of Biomolecular Structure and Dynamics, 2015:33(10):2221-2233.
13. Wu Z, Cheng F, Li J, Li W, Liu G, Tang Y. SDTNBI: An integrated network and chemoinformatics tool for systematic prediction of drug–target interactions and drug repositioning. Briefings in Bioinformatics, 2016, bbw012.
14. Cheng F, Zhou Y, Li W, Liu G, Tang Y. Prediction of Chemical-Protein Interactions Network with Weighted Network-Based Inference Method. PLoS ONE,2012:7(7):e41064.
15. 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 Computational Biology,2012:8(5):e1002503.
16. Cheng F, Li W, Wu Z, Wang X, Zhang C, Li J *et al*. Prediction of polypharmacological profiles of drugs by the integration of chemical, side effect, and therapeutic space. Journal of chemical information and modeling,2013:53(4):753-62.
17. Lü L, Medo M, Yeung CH, Zhang YC, Zhang ZK, Zhou T. Recommender systems. Physics reports, 2012:519(1):1-49.
18. Pan L, Zhou T, Lü L, Hu CK. Predicting missing links and identifying spurious links via likelihood analysis. Scientific reports,2016:6(1):1-0.
19. 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(12):i18-27.
20. Wu Z, Li W, Liu G, Tang Y. Network-based methods for prediction of drug-target interactions. Frontiers in pharmacology,2018:9:1134.
21. Durán C, Daminelli S, Thomas JM, Haupt VJ, Schroeder M, Cannistraci CV. Pioneering topological methods for network-based drug–target prediction by exploiting a brain-network self-organization theory. Briefings in bioinformatics,2018:19(6):1183-202.
22. Wu Z, Lu W, Yu W, Wang T, Li W, Liu G *et al*. Quantitative and systems pharmacology 2. In silico polypharmacology of G protein-coupled receptor ligands via network-based approaches. Pharmacological research,2018:129:400-13.
23. Clauset A, Moore C, Newman ME. Hierarchical structure and the prediction of missing links in networks. Nature,2008:453(7191):98-101.
24. Guimerà R, Sales-Pardo M. Missing and spurious interactions and the reconstruction of complex networks. Proceedings of the National Academy of Sciences,2009:106(52):22073-8.
25. Brown AS, Patel CJ. A standard database for drug repositioning. Scientific data,2017:4(1):1-7.
26. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ *et al*. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science,2006:313(5795):1929-35.
27. 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.
28. Li T, Wernersson R, Hansen RB, Horn H, Mercer J, Slodkowicz G *et al*. A scored human protein–protein interaction network to catalyze genomic interpretation. Nature methods,2017:14(1):61.
29. Zarin DA, Tse T, Williams RJ, Califf RM, Ide NC. The Clinical Trials. gov results database—update and key issues. New England Journal of Medicine, 2011:364(9):852-60.
30. Schork NJ. Artificial intelligence and personalized medicine. Precision Medicine in Cancer Therapy. Springer, Cham, 2019, 265-283.
31. Ginsburg GS, McCarthy JJ. Personalized medicine: revolutionizing drug discovery and patient care. Trends in Biotechnology,2001:19(12):491-6.