SVN Stroke and Vascular Neurology

Artificial intelligence and big data facilitated targeted drug discovery

Benquan Liu, Huiqin He, Hongyi Luo, Tingting Zhang, Jingwei Jiang

To cite: Liu B, He H, Luo H, *et al*. Artificial intelligence and big data facilitated targeted drug discovery. *Stroke & Vascular Neurology* 2019;**4**: e000290. doi:10.1136/svn-2019-000290

Received 16 October 2019
Accepted 28 October 2019
Published Online First
7 November 2019

Jiangsu Key Lab of Drug Screening, China Pharmaceutical University, Nanjing, China

Correspondence to
Dr Jingwei Jiang;
jiangjingwei@cpu.edu.cn

ABSTRACT

Different kinds of biological databases publicly available nowadays provide us a goldmine of multidiscipline big data. The Cancer Genome Atlas is a cancer database including detailed information of many patients with cancer. DrugBank is a database including detailed information of approved, investigational and withdrawn drugs, as well as other nutraceutical and metabolite structures. PubChem is a chemical compound database including all commercially available compounds as well as other synthesisable compounds. Protein Data Bank is a crystal structure database including X-ray, cryo-EM and nuclear magnetic resonance protein three-dimensional structures as well as their ligands. On the other hand, artificial intelligence (AI) is playing an important role in the drug discovery progress. The integration of such big data and AI is making a great difference in the discovery of novel targeted drug. In this review, we focus on the currently available advanced methods for the discovery of highly effective lead compounds with great absorption, distribution, metabolism, excretion and toxicity properties.

INTRODUCTION

Traditionally, the discovery of novel targeted drugs is an expensive long-term progress, costing billions of US dollars and more than 10 years. In the very beginning, a therapeutic drug target must be identified by traditional experimental methods. Then, structural biologists come to decipher the three-dimensional (3D) structures as well as their ligand-binding characteristics to reveal whether this is a druggable target. Subsequently, medicinal chemists and pharmacologists use high-throughput screening to find several highly effective lead compounds for further safety assessment as well as clinical trials. In general, the above procedures are costly and tedious. In November 2018, a study was conducted to estimate the total cost of trials for the development of novel Food and Drug Administration (FDA)-approved drugs. Surprisingly, the average cost of efficacy trials for the 59 new drugs approved by the FDA during 2015–2016 was $19 million.[1] Therefore, it is necessary to overcome the limitations of the conventional drug discovery procedures by introducing efficient, low-cost and computational methods.

Compared with traditional drug discovery methods, rational drug design, mainly including computer-aided drug design (CADD), is more efficient and economical. Rational drug design integrates molecular docking to the ligand-binding pocket of a promising therapeutic target, computes the binding energy of each docked small molecule compound, and selectively chooses the best ones as candidates for subsequent experimental procedures. Today, there are more than 100 000 protein 3D structures deposited in Protein Data Bank (PDB) for molecular docking.[2] In contrast to traditional methods, rational drug design has boosted the hit rate of drug screening by more than 100 times, from ~0.01% to 1%~2%. Moreover, CADD is a more multidiscipline method which integrates advanced bioinformatic techniques and sophisticated computational algorithms. Due to its relatively high hit rates, CADD method is becoming the fundamental basis of industrial drug discovery as well as academic research.[3] Recently, artificial intelligence (AI) assisted drug discovery of me-better drug from hit discovery to animal tests within 46 days.[4]

Cancer-targeted drugs are the most successful drugs for the last three decades, thanks to comprehensive omics databases of cancer research. A lot of cancer-related proteins have been identified as therapeutic targets by computational data mining of transcriptome data in databases such as The Cancer Genome Atlas (TCGA),[5] The Human Protein Atlas(THPA)[6] and so on. Unfortunately for other diseases, such as stroke, vascular-related diseases and other genetic diseases, there are no similar integrated omics databases to provide sufficient big data. However, there are increasingly more single cell transcriptome data of various diseases publicly available.[7–10] Thus, such data will be precious goldmines in terms of the discovery of therapeutic targets for stroke, vascular-related diseases and other genetic diseases. Moreover, supercomputers are speeding up lead identification and evaluation. In this review, we provide an overview of how the integration of big data and AI could help us to discover new therapeutic targets and their targeted lead compounds, as well as

their absorption, distribution, metabolism, excretion and toxicity (ADMET) properties.

VIRTUAL SCREENING TO DISCOVER TARGETED LEAD COMPOUNDS

Virtual screening technology is the core of CADD. Based on the 3D structure or the quantitative structure–activity relationship model of the target biomacromolecules, the theory of molecular biology and computer science and other related fields is used as a technical basis to select the compounds that meet the expectations from the known small-molecule databases. Then, one or more experimental methods are selected for targeted drug screening for specific diseases. In the pharmaceutical world, virtual screening is often considered as a top CADD tool to screen large chemical structural libraries and reduce them to a set of candidate compounds related to specific protein targets.[11] At present, virtual screening has been regarded as a materialised tool, widely recognised in search for lead compounds and the enhancement of compound activity.[12]

The basic processes of virtual filtering mainly include the following:

- Target selection: this is the first step in virtual screening, and this step is crucial. Small molecular compounds target four large molecules: proteins, polysaccharides, lipids and nucleic acids. Proteins such as enzymes, ion channels and GPCR(G Protein-coupled Receptor) are often preferred as potential drug targets because they are highly specific and less toxic, such as the discovery of heat shock protein (Hsp90) inhibitors,[13] the discovery of a selective inhibitor of Aurora A,[14] the discovery of TASK-3 (KCNK9) channel blockers,[15] the virtual screening for GPCR drug screening[16] and so on.
- Prepare the compound database: before starting a new virtual screening, we need to collect all the compound structures for a specific drug target. In recent years, a number of compound databases have been developed which store not only the structure of the compound molecules, but also many chemical and biological information, such as ZINC,[17] PubChem[18] and others.
- Docking software: currently popular molecular docking software are Dock, AutoDock, MolDock, Maestro and so on.[19] These software are available for use and are easy to operate, but when the number of compounds involved in docking is too large (eg, 1 million), large-scale molecular docking methods and strategies need to be adopted. Linux-based virtual docking always plays an important role when we perform high-throughput docking.
- Scoring system: molecular docking is a computational method that predicts the preferred position of a molecule (ligand) relative to a second molecule (receptor) when the two molecules combine to form a stable complex, and then predict the binding strength or binding affinity between the receptor and the ligand. There are two main types of docking: rigid docking and flexible docking.[20] In rigid docking, the receptor and ligand are immobilised so that the bond angle and bond length are constant. This docking speed is very fast, but lacks practical application because flexible docking allows for conformational transformation. In flexible docking, the conformation of the ligand and acceptor can be converted at will during the calculation. This docking method requires relatively high computing power, but it can most accurately calculate the docking results and is suitable for the accurate investigation of the identification between molecules. Based on the position and binding energy, a docking score will be calculated.
- Biological experiment verification: the candidate compounds of highest docking score are verified by both in vitro and in vivo biological experiments.
- Clinical study: once all preclinical studies of these candidate compounds are proved to be effective, clinical studies will be performed on candidate compounds to determine their safety and effectiveness on patients.

IDENTIFICATION OF LIGAND-BINDING POCKET ON THE 3D PROTEIN MODEL

The interaction between protein and ligand usually occurs in a pocket formed by conserved amino acids. The protein function relies on the ligand-binding site on its 3D structure. The identification of the binding pocket helps to discover new drugs and better understand the mechanism of actions of drugs, such as the discovery of a conservative pocket of the guanylate cyclase heme domain.[21] In the general molecular docking calculation, an indispensable step is to define the binding position of the ligand molecule, that is, its binding pocket. If the binding site is known, the ligand type and protein function can be determined by computer and experimental procedures, and can be used in drug design and to predict potential side effects.[22]

Bioinformatics is a cross-disciplinary discipline that solves biological problems through the use of computer, mathematical and statistical methods. The determination of binding pockets is very important for designing drugs. Traditional X-ray crystallography and nuclear magnetic resonance methods predict large amounts of protein structures that are time-consuming and expensive, but bioinformatics provides different tools to predict the 3D structure of proteins and reveal their binding regions. Its application is very promising, such as the identification of conserved binding pockets in ricin A chain,[23] RASSF2 potential binding pocket prediction[24] and so on. There are two ways to find a pocket combination: (1) proteins with known 3D structures can be searched from the PDB database,[25] and related information can be downloaded directly from the database; and (2) method of homology modelling, using I-TASSER, SwissModel, ModWeb and other online servers based on homologous modelling to

generate protein 3D structure, as well as to predict the ligand-binding pocket, for example, prediction of serotonin 1A receptor binding pocket.[26]

DISCOVERY OF TARGETED LEAD COMPOUNDS FOR A NOVEL DRUG TARGET

The drug target is a special site formed by biomolecules, and the drug can be combined with it to produce pharmacological effects (targeted agonist/inhibitor) for the purpose of preventing and treating diseases. According to the biological characteristics of biomolecules, drug targets can be classified into receptors, enzymes, ion channels, DNA, hormones and growth factors. The research and development (R&D) of new drug is a work with high investment and low yield. The discovery and confirmation of drug target is the first step of the R&D of a new drug. However, the number of clinically validated drug targets is still very small, so there is an urgent need to discover more new drug targets.

With the development of life science and bioinformatics, more and more target structures have been analysed. Different from traditional drug research methods, big data mining is widely used in drug target research, such as using genetic algorithm and bagging-svm ensemble classifier to predict drug targets,[27] mining and forecasting cancer-related database,[28] and using genetic disease-related data to predict novel therapeutic targets by computational data mining methods.[29]

The human genome database shows that there are more than 20 000 proteins in the human body, while the DrugBank database indicates only about 500 have been identified in the past 100 years.[30] Therefore, there are many potential targets to be discovered and confirmed. Thanks to structure biologists, a lot of new biological processes mediated by protein–protein interaction, protein–DNA interaction and protein–RNA interaction have been discovered. These above proteins may probably serve as potential novel drug targets in the near future. The information of the drug target database can be used to analyse the sequence characteristics and biochemical characteristics of structural features, and to establish a prediction model to discover new drug targets. Therefore, we set up a set of novel methods for potential cancer-related drug target discovery, such as the following procedures: (1) TCGA and Human Protein Atlas databases were used to mine the data of targets related to prediction of cancer prognosis in the database. (2) Then use the computer to correlate with known cancer prognosis-related targets and score according to the correlation strength. (3) Then review the research progress of the target according to the score table and explore the 3D structural information of the drug target in the PDB database. (4) According to the integrated information, select the appropriate targets for further biological verification.

After successful verification of the novel therapeutic targets in vitro and in vivo, the virtual screening molecular docking-based drug screening can be performed according to the novel targets. This process has greatly reduced the time and cost compared with traditional drug development. In the past few years, our lab has discovered 73 novel compounds as well as 12 FDA-approved drugs targeting more than 30 potential novel therapeutic targets (figure 1). Moreover, four FDA-approved drugs will be used for clinical trial tests to cure multiple sclerosis in the near future (Drug repositioning, unpublished data,Jingwei Jiang).

REVERSE DOCKING TO FIND DRUG TARGETS OF AN OLD DRUG

Drug repositioning, also known as drug repurposing, defines new indications for existing drugs and can be used as an alternative to drug development.[31] The benefits of repositioning include the availability of chemical materials and previously generated data that can be used, so the potential for R&D is significantly greater than the time and cost-effectiveness of bringing new drugs to market. It has been reported that there have been identified 109 molecules with other activities through in vitro screening, with these products having at least one marketing approval for a common disease indication or one marketing approval for a rare disease from the FDA's rare disease research database.[32] In our meta-analysis, a study shows that the class III antiarrhythmic amiodarone was active in neurodegeneration assays and could also selectively remove embryonic stem cells, and that the antipsychotic trifluoperazine was active in neurodegeneration assays.[32] In contrast to traditional molecular docking, reverse docking is used for identifying receptors for a given ligand among a large number of crystal structures. It can be used to discover new targets for existing drugs and natural compounds, alternative indications of drugs through drug repositioning, and detecting adverse drug reactions and drug toxicity.[33] Generally, the following steps are required to perform a drug repositioning by reverse docking (drug repositioning): (1) data set collection; (2) data set partition; (3) molecular descriptor calculation and modelling; (4) ensemble learning; (5) retrospective screening campaigns; (6) building positivity predictive value surfaces and choosing an adequate score threshold value; (7) prospective virtual screening; (8) molecular docking; and (9) reverse docking scoring. The results of the reverse docking were then verified by biological experiments. There are reports that they have implemented a computer-aided drug repurposing campaign to discover new inhibitors of falcipain-2. Four hits were acquired and tested against the enzyme, with two of them confirming inhibitory activity.[34] The abandoned drug odanacatib displayed competitive inhibition, while the antibiotic methacycline also showed inhibitory effects through non-competitive inhibition.[34] Therefore, it is feasible to find the target of the old drug through reverse docking. This method saves a lot of time and can reduce many experimental costs and experimental steps.

In the past few years, our lab has discovered 13 new targets for eight FDA-approved drugs through

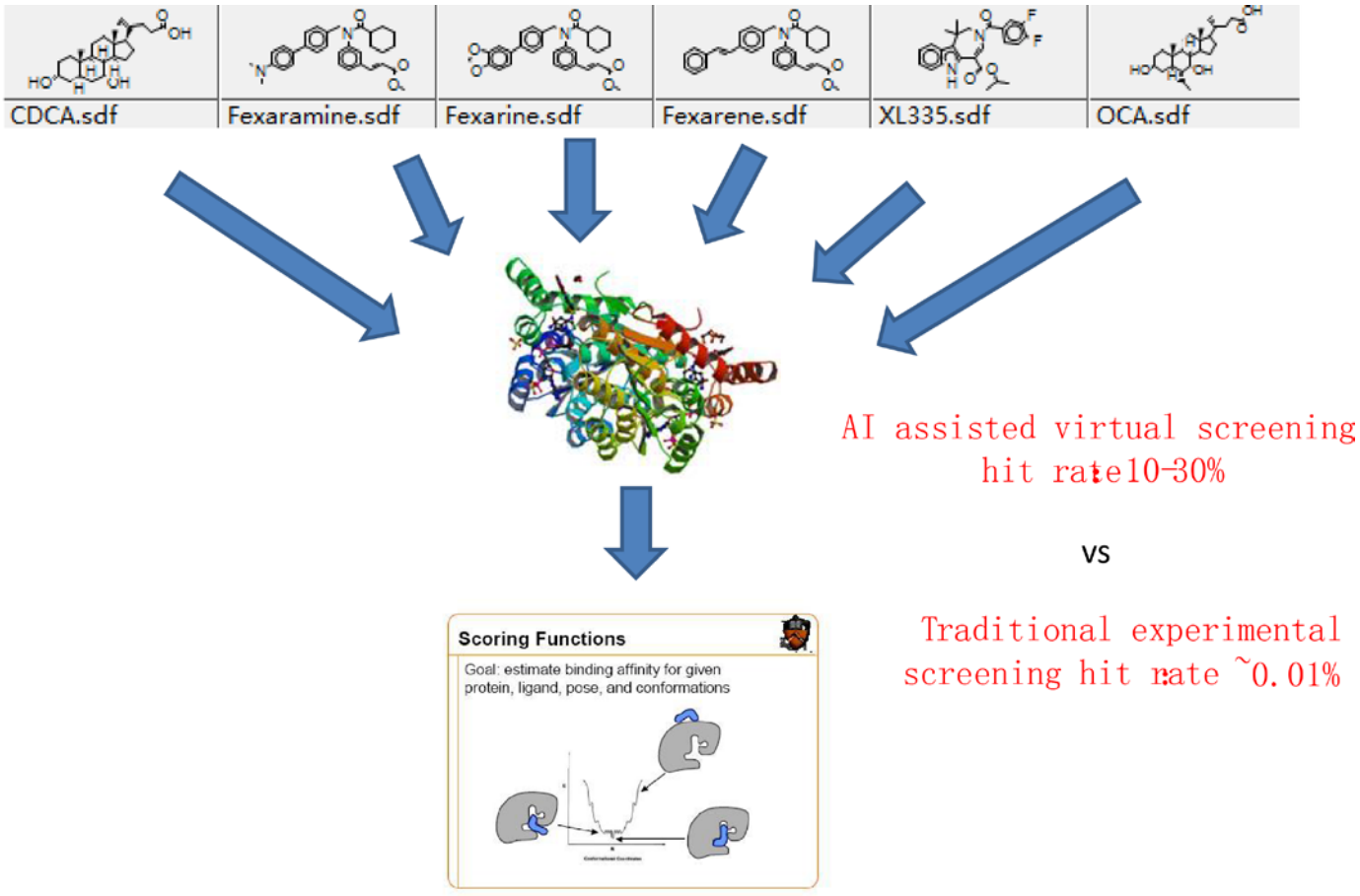

Figure 1 Schematic procedure of artificial intelligence (AI)-assisted virtual screening. Millions of structurally diverse chemical compounds are docked to a specific therapeutic target. AI scoring function is used to select the best hits from millions of docked results.

reverse-docking the old drugs to all ligand-binded structures extracted from PDB (>100 000 proteins; figure 2). The new indications and adverse effects of these old drugs have been revealed through biological verification for those reverse-docked targets (unpublished data,Jingwei Jiang).

AI FOR THE PREDICTION OF A COMPOUND'S ADMET

The ADMET of chemicals plays a key role in drug discovery and development. High-quality drug candidates should not only have sufficient efficacy for the treatment target, but should also display appropriate ADMET characteristics at the therapeutic dose.[35 36] Moreover, ADMET's predictions not only reduce the risk of late-stage attrition of new compounds and compound libraries but also help researchers optimise screening and testing by looking at only the most promising compounds.[37] Just relying on biological experiments to verify the ADMET of a compound is a waste not only of time but also a lot of human and material resources. With the increase in computer speed and the implementation of quantum chemistry methodology, pharmacodynamic and pharmacokinetic issues have become computationally easier to handle. Quantum mechanics provides pharmaceutical scientists with the opportunity to study pharmacokinetic problems at the molecular level prior to laboratory preparation and testing.[38] In order to realise ADMET for predicting compounds by computer, we need to do a lot of work in the early stage: (1) data collection and preparation (this is a crucial step); (2) calculation of ADMET-related properties based on the collected data; (3) definition of the ADMET score; and (4) validation of the ADMET score.[36] An article on predicting the antimalarial activity of artemisinin derivatives showed that their predicted results showed significant antimalarial activity of compounds A24, A24a, A53, A54, A62 and A64. Subsequent studies of the derivative A64 showed that the experimental results of the derivative were well agreed with the predicted values.[39] Although it is not guaranteed that the predicted results are completely consistent with the later experimental results, the introduction of AI can reduce many unnecessary troubles for later research. Machine learning (including AI) methods are accompanied by verification procedures in many cases and are often used in conjunction with other methods. Therefore, this makes them an excellent and attractive hybrid tool for reducing false predictions and model errors.[40]

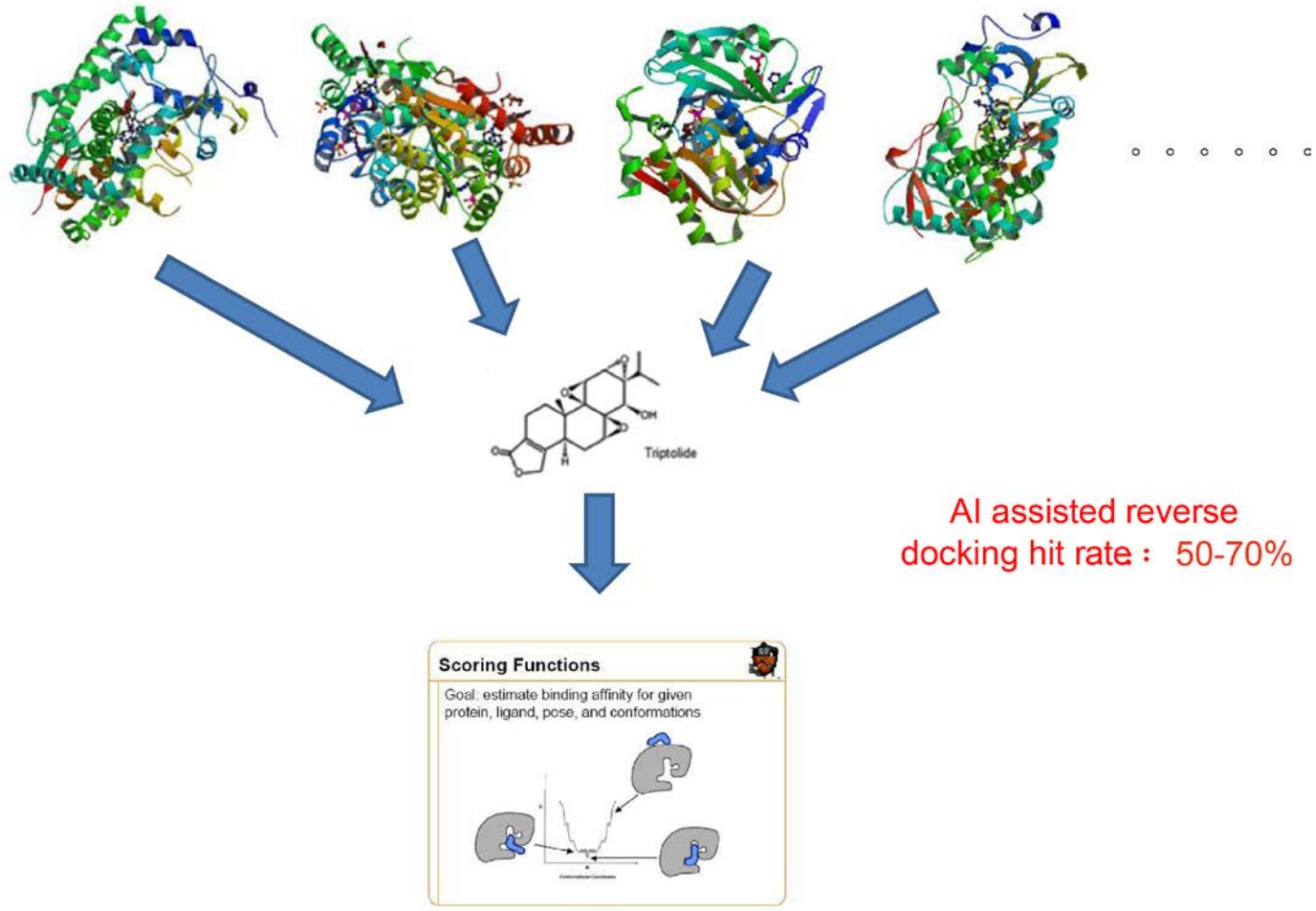

Figure 2 Schematic procedure of artificial intelligence (AI)-assisted reverse docking. More than 100 000 structurally diverse protein structures are reversely docked to a specific chemical compound/natural product. AI scoring function is used to select the best hits from millions of docked results.

In the past few years, our lab has developed several new ADMET prediction tools based on deep learning AI, such as prediction for logBB and logPapp to calculate the overall ADMET properties of a specific compound (figure 3A,B). Toxicity of a compound is very difficult to predict, mainly because it depends not only on its own chemical structure but also its direct actions on the target proteins. Hence, we collected all FDA-approved/withdrawn drugs (June 2019) to perform batch reverse docking with all ligand-binded structures extracted from PDB. Every docked target of each drug was scored and each drug can be considered as an N-dimensional vector in an N-dimensional space (figure 3C,D). Therefore, FDA-approved/withdrawn drugs can be referred as training data set to predict the toxicity of any given compounds (unpublished data,Jingwei Jiang).

MINING CANCER DATABASE TO DISCOVER NOVEL THERAPEUTIC DRUG TARGETS

Targeted drug design has become a hot topic because it is one of the key technologies for the discovery of therapeutic drugs. However, it is very difficult to find new drug targets through traditional experiments and methods and it is often difficult to achieve the desired results. Therefore, bioinformatic technology can be used to discover and identify new drug targets by mining cancer database. With the complete information of cancer genome/transcriptome sequencing accumulated in recent years, a variety of publicly available biological databases have provided us with a multidisciplinary goldmine of big data; especially the cancer genomic/transcriptomic/proteomic research has taken a big step forward.

TCGA is a project jointly supervised by the National Cancer Institute and the National Human Genome Research Institute. It aims to use high-throughput genome analysis technology to help people to better understand the occurrence and development of cancer, in order to achieve the purpose of prevention, diagnosis and treatment.[41] For example, as of 2012, the genomes and epigenetic groups of lung squamous cell carcinoma have not been fully elucidated, but through the genomic and epigenetic analyses of about 180 lung SQCCs (Squamous Cell Carcinoma), TCGA has successfully screened out molecular targeting drugs for SQCC.[42] Similarly,

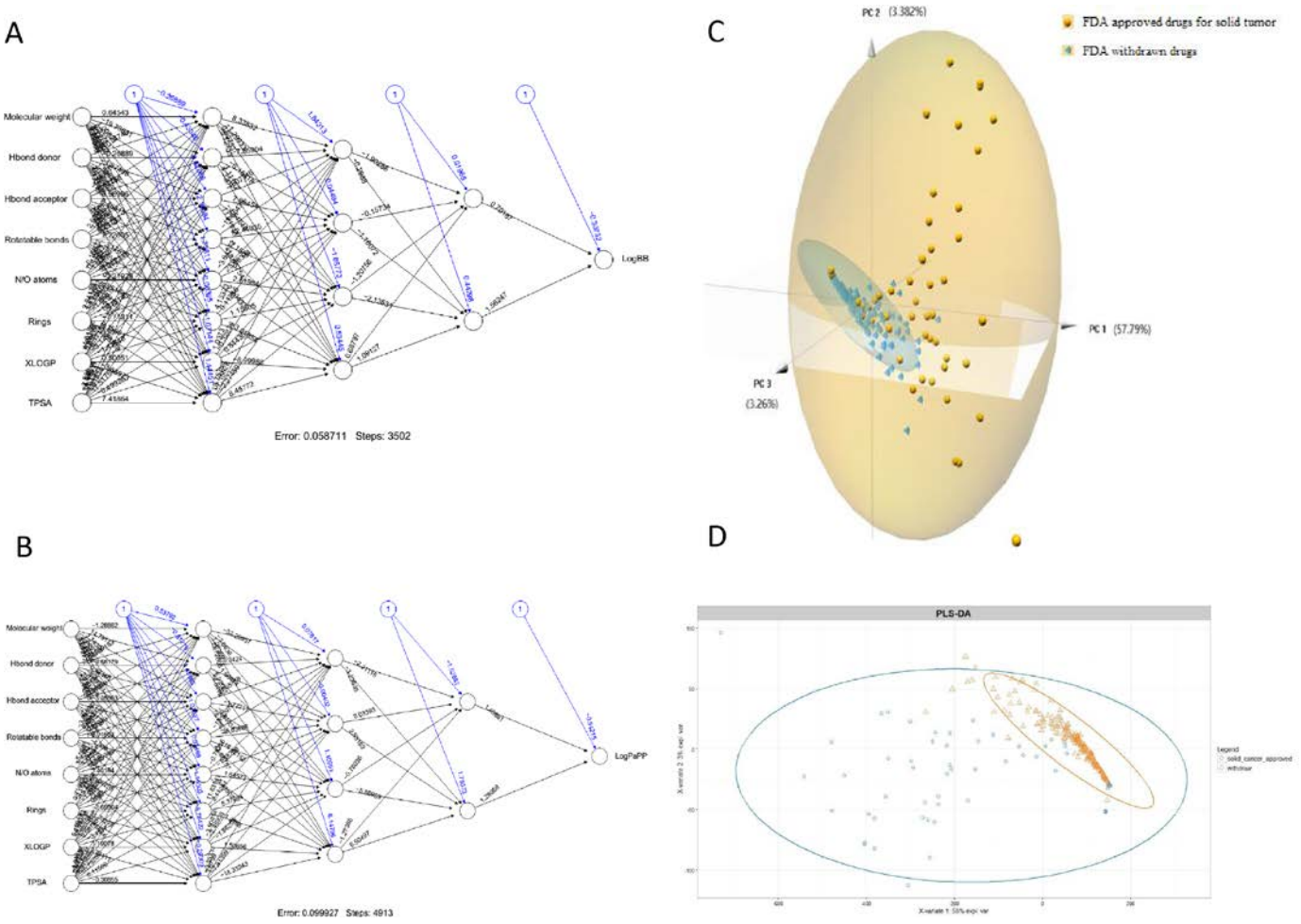

Figure 3 AI-assisted ADMET properties prediction. (A) Deep learning algorithm to calculate logBB for a specific chemical compound. (B) Deep learning algorithm to calculate logPapp for a specific chemical compound. (C) PCA(Principal Component Analysis) analysis on 48 186 reverse-docked proteins for 55 FDA-approved drugs (yellow dots) and 224 FDA-withdrawn drugs (blue dots). (D) PLS-DA(Partial Least Squares Discriminant Analysis) analysis on 48 186 reverse-docked proteins for 55 FDA-approved drugs (blue dots) and 224 FDA-withdrawn drugs (yellow dots). ADMET, absorption, distribution, metabolism, excretion and toxicity; AI, artificial intelligence; FDA, Food and Drug Administration;TPSA,total polar surface area.

for ovarian serous cystadenocarcinoma, which is not optimistic in diagnosis and treatment at present, some potential therapeutic targets have been found through the comprehensive analysis of ovarian serous cystadenocarcinoma with higher grade by TCGA.[43] In this way, the mining of the cancer database plays an important role in finding new therapeutic druggable targets.

By mining TCGA and THPA, our lab has discovered more than 10 potential novel therapeutic targets in various cancers, such as pancreatic cancer, lung cancer, triple negative breast cancer, colorectal cancer and so on, as well as their targeted compounds recently (unpublished data,Jingwei Jiang). For other diseases (such as stroke, cardiovascular diseases, neurological diseases and so on), there is no such intact database for the data mining to discover novel therapeutic targets. However, single cell transcriptomic sequencing data have been accumulated rapidly in the recent years, and these data will be helpful for new therapeutic target discovery in the near future.

8.ANIMAL MODELS AND THEIR LIMITATIONS

In order to study the physiological and biochemical processes of human diseases and to explore the pharmacodynamics and pharmacokinetics of drugs in vivo, many animal models of various diseases have been introduced to preclinical studies. The most popular animal models are mouse, rat and monkey. Particularly, specific genes knock-out/knock-in mouse models have revolutionised our ability to study specific gene and protein functions in vivo and to better understand their molecular pathways and mechanisms.[44]

Although there are animal models used as powerful support for modern medical research in preclinical studies of many diseases, the new drug therapy is still difficult to convert from laboratory to clinical, because it is not feasible to mimic all aspects of a human disease in an animal model, especially a heterogeneous disease with complex pathophysiology such as stroke, and most of its studies are carried out in young animals without any

complications. These models are physiologically different from real stroke, which especially affects the elderly with a variety of cerebrovascular risk factors.[45] Therefore, in stroke studies, more than 1000 drugs were candidates in stroke models, but only 17 were tested in humans.[46] Recently, many Alzheimer's disease (AD) candidate drugs have shown great effects in mouse models but all failed during clinical trials. Perhaps this is because the tissues, organs and systems of animals are always different from those of human beings, and their reactions and effects to drugs are also different. An animal model cannot involve all aspects of a human disease. The age, sex and species of animals, tissue and organ damage, or the increase, deletion and change of genes caused by the establishment of animal models may have a significant impact on the experimental results.

Furthermore, another big problem of an animal model is the genetic difference between the animal protein and human protein. According to ENSEMBL genome database, orthologous genes have been analysed in human, chimpanzee, mouse and rat. Surprisingly, there are only 7043 orthologous genes (single copy common genes) shared in these four species. For chimpanzee and human, a set of 13 454 pairs of human and chimpanzee genes with unambiguous 1:1 orthology have been identified. Orthologous proteins in human and chimpanzee are extremely similar, with ~29% being identical and the typical orthologue differing by only two amino acids, one per lineage.[47] Compared with ~25 000 genes in each of these four species, 7043 orthologous genes are ~28%, which means the other ~72% expressed non-orthologous proteins in these four species are very different in their protein sequences. Even if humans and chimpanzees are considered as the closest primate relatives in the animal kingdom, only 13 454 pairs of orthologue genes are identified consisting ~50% of their own expressed genes, which means the other ~50% expressed non-orthologous proteins are very different in their amino acid sequences. Taken together this above genetic evidence, it is very clear that if the drug target of the animal model is structurally different from the one of human, drugs targeting the animal protein will perform a significantly different effect between animal experiment and clinical experiment. The interaction between drug and its target is caused by hydrogen bonds, Van der Waals force and ϖ-ϖ interaction, which are exerting their interactive forces within less than 4 Angstrom. One or two amino acid mutations within the binding pocket of the drug target can make a big difference.

Cancer-targeted drugs are much more successful compared with targeted drugs developing for stroke and AD. Perhaps there are two major reasons. First, in the field of cancer-targeted drug R&D, there are a lot of mouse models carrying humanised genes (such as mouse carrying humanised immune system) to mimic the human immunity system. Second, patient-derived xenograft models (mouse carrying clinical human cancer tissue) have been widely introduced in the preclinical studies of cancer-targeted drugs. For stroke, AD and rare diseases, similar humanised animal models carrying human drug target protein must also be introduced in the preclinical studies in the near future.

CONCLUSION REMARK

Today, big data and AI are developing so fast that boost targeted drug discovery in an unprecedented speed. With the integration of various disease databases, scientists are able to perform data mining for de novo therapeutic target discovery. With AI assistance, novel identified therapeutic targets can be virtually screened for the discovery of targeted old drugs/new compounds within very short period. With AI-assisted reverse docking, old drugs or natural products could be repurposed for new indications very efficiently. ADMET properties can also be predicted by AI deep learning models to boost the success rate of in vivo experiments. Finally, with the help of 3D therapeutic target structural alignment, scientists can identify the difference on the drug binding pocket of a specific therapeutic target between human and animal models, and the selection of animal model must be considered very carefully in terms of their target 3D similarity.

Contributors All authors wrote the manuscript. JJ provided guidance and modifications.

Funding This work was supported by NSFC (no 81872892 and no 2018ZX09735001-004) and 'Double First Class' University project (no CPU2018GY20 and no CPU2018GY38).

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Commissioned; externally peer reviewed.

REFERENCES

1 Moore TJ, Zhang H, Anderson G, *et al*. Estimated costs of pivotal trials for novel therapeutic agents Approved by the US food and drug administration, 2015-2016. *JAMA Intern Med* 2018;178:1451–7.
2 Ferreira LG, Dos Santos RN, Oliva G, *et al*. Molecular docking and structure-based drug design strategies. *Molecules* 2015;20:13384–421.
3 da Silva Rocha SFL, Olanda CG, Fokoue HH, *et al*. Virtual screening techniques in drug discovery: review and recent applications. *Curr Top Med Chem* 2019;19:1751–67.
4 Zhavoronkov A, Ivanenkov YA, Aliper A, *et al*. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nat Biotechnol* 2019;37:1038–40.
5 Tomczak K, Czerwińska P, Wiznerowicz M. The cancer genome atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol* 2015;19:A68–77.
6 Colwill K, Gräslund S, Renewable Protein Binder Working Group. A roadmap to generate renewable protein binders to the human proteome. *Nat Methods* 2011;8:551–8.
7 Gupta I, Collier PG, Haase B, *et al*. Single-Cell isoform RNA sequencing characterizes isoforms in thousands of cerebellar cells. *Nat Biotechnol* 2018;36:1197–202.
8 Gladka MM, Molenaar B, de Ruiter H, *et al*. Single-Cell sequencing of the healthy and diseased heart reveals cytoskeleton-associated protein 4 as a new modulator of fibroblasts activation. *Circulation* 2018;138:166–80.

9 Kaur H, Carvalho J, Looso M, *et al*. Single-Cell profiling reveals heterogeneity and functional patterning of GPCR expression in the vascular system. *Nat Commun* 2017;8:15700.
10 Liu X, Chen W, Li W, *et al*. Single-Cell RNA-seq of the developing cardiac outflow tract reveals convergent development of the vascular smooth muscle cells. *Cell Rep* 2019;28:1346–61.
11 Cerqueira NMFSA, Gesto D, Oliveira EF, *et al*. Receptor-Based virtual screening protocol for drug discovery. *Arch Biochem Biophys* 2015;582:56–67.
12 Ferreira L, dos Santos R, Oliva G, *et al*. Molecular docking and structure-based drug design strategies. *Molecules* 2015;20:13384–421.
13 Abbasi M, Amanlou M, Aghaei M, *et al*. New heat shock protein (Hsp90) inhibitors, designed by pharmacophore modeling and virtual screening: synthesis, biological evaluation and molecular dynamics studies. *J Biomol Struct Dyn* 2019;5:1–12.
14 Kilchmann F, Marcaida MJ, Kotak S, *et al*. Discovery of a selective Aurora a kinase inhibitor by virtual screening. *J Med Chem* 2016;59:7188–211.
15 Ramírez D, Concha G, Arévalo B, *et al*. Discovery of novel TASK-3 channel blockers using a pharmacophore-based virtual screening. *Int J Mol Sci* 2019;20:4014.
16 Chen J-Z, Wang J, Xie X-Q. Gpcr structure-based virtual screening approach for CB2 antagonist search. *J Chem Inf Model* 2007;47:1626–37.
17 Irwin JJ, Sterling T, Mysinger MM, *et al*. Zinc: a free tool to discover chemistry for biology. *J Chem Inf Model* 2012;52:1757–68.
18 Kim S, Chen J, Cheng T, *et al*. PubChem 2019 update: improved access to chemical data. *Nucleic Acids Res* 2019;47:D1102–9.
19 Gupta M, Sharma R, Kumar A. Docking techniques in pharmacology: how much promising? *Comput Biol Chem* 2018;76:210–7.
20 Forli S, Huey R, Pique ME, *et al*. Computational protein–ligand docking and virtual drug screening with the AutoDock suite. *Nat Protoc* 2016;11:905–19.
21 Wales JA, Chen C-Y, Breci L, *et al*. Discovery of stimulator binding to a conserved pocket in the heme domain of soluble guanylyl cyclase. *J Biol Chem* 2018;293:1850–64.
22 Hoffmann B, Zaslavskiy M, Vert J-P, *et al*. A new protein binding pocket similarity measure based on comparison of clouds of atoms in 3D: application to ligand prediction. *BMC Bioinformatics* 2010;11:99.
23 Zhou CLE, Zemla AT, Roe D, *et al*. Computational approaches for identification of conserved/unique binding pockets in the A chain of ricin. *Bioinformatics* 2005;21:3089–96.
24 Kanwal S, Jamil F, Ali A, *et al*. Comparative modeling, molecular docking, and revealing of potential binding pockets of RASSF2; a candidate cancer gene. *Interdiscip Sci* 2017;9:214–23.
25 Desaphy J, Bret G, Rognan D, *et al*. sc-PDB: a 3D-database of ligandable binding sites—10 years on. *Nucleic Acids Res* 2015;43:D399–404.
26 Warszycki D, Rueda M, Mordalski S, *et al*. From Homology Models to a Set of Predictive Binding Pockets-a 5-HT_{1A} Receptor Case Study. *J Chem Inf Model* 2017;57:311–21.
27 Lin J, Chen H, Li S, *et al*. Accurate prediction of potential druggable proteins based on genetic algorithm and Bagging-SVM ensemble classifier. *Artif Intell Med* 2019;98:35–47.
28 Cova TFGG, Bento DJ, Nunes SCC. Computational approaches in theranostics: mining and predicting cancer data. *Pharmaceutics* 2019;11:119.
29 Ferrero E, Dunham I, Sanseau P. In silico prediction of novel therapeutic targets using gene–disease association data. *J Transl Med* 2017;15:182.
30 Yildirim MA, Goh K-I, Cusick ME, *et al*. Drug-target network. *Nat Biotechnol* 2007;25:1119–26.
31 Kim E, Choi A-sol, Nam H. Drug repositioning of herbal compounds via a machine-learning approach. *BMC Bioinformatics* 2019;20:247.
32 Ekins S, Williams AJ. Finding promiscuous old drugs for new uses. *Pharm Res* 2011;28:1785–91.
33 Lee A, Lee K, Kim D. Using reverse docking for target identification and its applications for drug discovery. *Expert Opin Drug Discov* 2016;11:707–15.
34 Alberca LN, Chuguransky SR, Álvarez CL, *et al*. *In silico* guided drug repurposing: discovery of new competitive and non-competitive inhibitors of falcipain-2. *Front Chem* 2019;7:534.
35 Hessler G, Baringhaus K-H. Artificial intelligence in drug design. *Molecules* 2018;23:E2520.
36 Guan L, Yang H, Cai Y, *et al*. ADMET-score - a comprehensive scoring function for evaluation of chemical drug-likeness. *Medchemcomm* 2019;10:148–57.
37 van de Waterbeemd H, Gifford E. Admet in silico modelling: towards prediction paradise? *Nat Rev Drug Discov* 2003;2:192–204.
38 Bowen JP, Güner OF, Osman FG. A perspective on quantum mechanics calculations in ADMET predictions. *Curr Top Med Chem* 2013;13:1257–72.
39 Qidwai T, Yadav DK, Khan F, *et al*. Qsar, docking and ADMET studies of artemisinin derivatives for antimalarial activity targeting plasmepsin II, a hemoglobin-degrading enzyme from P. falciparum. *Curr Pharm Des* 2012;18:6133–54.
40 Dobchev DA, Pillai GG, Karelson M. In silico machine learning methods in drug development. *Curr Top Med Chem* 2014;14:1913–22.
41 Chin L, Andersen JN, Futreal PA. Cancer genomics: from discovery science to personalized medicine. *Nat Med* 2011;17:297–303.
42 Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* 2012;489:519–25.
43 Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. *Nature* 2011;474:609–15.
44 Rathore K, Cekanova M. Animal model of naturally occurring bladder cancer: characterization of four new canine transitional cell carcinoma cell lines. *BMC Cancer* 2014;14:465.
45 Fluri F, Schuhmann MK, Kleinschnitz C. Animal models of ischemic stroke and their application in clinical research. *Drug Des Devel Ther* 2015;9:3445–54.
46 O'Collins VE, Macleod MR, Donnan GA, *et al*. 1,026 experimental treatments in acute stroke. *Ann Neurol* 2006;59:467–77.
47 Waterson RH, Lander ES, Wilson RK, *et al*. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 2005;437:69–87.