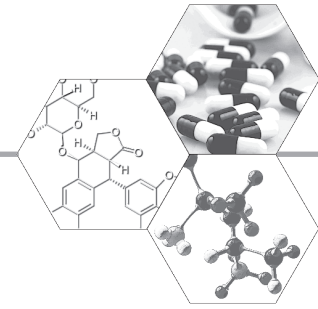


Computational target fishing: what should chemogenomics researchers expect for the future of *in silico* drug design and discovery?

"*Advances in cell signaling, chemical genomics and proteomics have provided new potential drug targets, whereas computational target fishing technologies increase our ability to efficiently and effectively screen against these targets in high-throughput formats.*"

Keywords: big data ■ chemogenomics ■ cloud computation ■ computational target fishing ■ information integration

Target fishing, or target identification, is an important start step in modern drug development, which investigates the mechanism of action of bioactive small molecules by identifying their interacting proteins. It can also be used to find potential off-targets of therapeutic compounds for the study of their side effects or for drug repurposing.

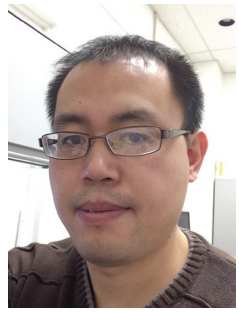
Computational target fishing employs chemoinformatic tools and machine learning algorithms for *in silico* prediction of the biological targets of a chemical. Much progress has been made in this field over the past 10 years and quite a few approaches have been developed, such as chemical structure similarity searching [1], data mining/machine learning [2], panel docking [3], and bioactivity spectra based algorithms [4], to name a few. For further reading, we would like to refer to the following reviews [5–7]. As computational target fishing can facilitate the quick identification of new drug targets, the prediction of possible off-targets to avoid adverse effects and even the evaluation of selectivity among protein families, such technologies can help to leverage the challenges faced of the pharmaceutical industry, such as a decreasing output of new drugs and increasing regulatory requirements of safety. However, targets and diseases now being addressed are increasingly complex and challenging. Efforts are needed to develop more powerful tools in order to accelerate the discovery of small-molecule modulators targeting novel protein classes. Here, we provide our perspectives on this field in the future.

Integration is the key

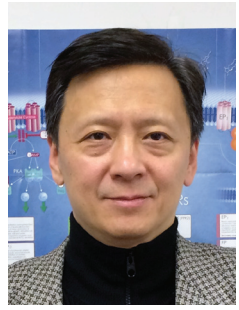
Computational or *in silico* target fishing has emerged as an interdisciplinary field with tremendous potential to advance *in silico* drug design and discovery. It merges the expertise of machine learning, chemoinformatics and bioinformatics technologies to develop approaches that can prioritize the possible targets of a chemical compound. Integrations play a key role in the development of approaches for computational target fishing. First of all, understanding fundamental biological mechanisms is essential for the development of drugs to manage diseases. In line with this, information integration is required to consider more bio-related data in different levels to make the prediction more reliable. Such data will comprise:

- Chemogenomics data, which include the binding affinity of small molecules against proteins. For example, ChEMBL [101] and PubChem [102] databases;
- Druggable target databases, such as potential drug target databases [103] and DrugBank [104];
- Therapeutic target database [105];
- PharmGKB [106] and DrugMap [107];
- Protein structure database, such as PDB [108];
- Protein expression information, from normal and different diseases, such as human protein atlas [109];
- Disease specific targets databases, such as MetaCore [110];
- Pathway information such as KEGG [111];
- Toxicity databases, such as the comparative toxicogenomics database [112] and the target-toxin database [113].

Here, we emphasize the importance of disease, cell type and tissue specific protein information. Such data will help us refine the ranking of target prediction by eliminating unrelated proteins and



Lirong Wang
Computational Chemical Genomics Screening Center, Department of Pharmaceutical Sciences, University of Pittsburgh, PA 15261, USA



Xiang-Qun Xie
Author for correspondence:
Computational Chemical Genomics Screening Center, Department of Pharmaceutical Sciences, School of Pharmacy, Drug Discovery Institute, University of Pittsburgh, Pittsburgh, PA 15261, USA
Tel.: +1 412 383 5276
Fax: +1 412 383 7436
xix15@pitt.edu

FUTURE SCIENCE part of fsg

pathways and focusing on the related parts. With the advance of genomics profiling via next-generation sequencing, computational target fishing can even be used for lead identification of personalized medicine by incorporating personal genomics information. On the other hand, the drug-likeness of small molecules will also need to be considered as the attrition will be reduced by incorporating drug property information in the early stages of drug development.

Approach integration is essential

To address different targets prediction tasks, combination of different complementary approaches have more advantages. Molecular fingerprints based similarity search is prevailing to find most similar analogs with their target annotated and to guide the test of the query compound against these proteins, whereas it is hard for target identification of compounds with novel scaffolds. Docking-based method relies on the availability of 3D structures of proteins. Machine learning approaches depend on reliable training data sets, and bioactivity spectra-based technologies rest on bio-profile data from experiments, which require huge resources, time and effort. By combining these approaches, we could overcome their drawbacks and achieve confident predications at different levels. We also need to highlight the importance of integration of traditional quantitative structure–activity/property relationship approaches, which will certainly be helpful for the prediction of polypharmacology effect of small molecules on multiple protein classes. Such trends have emerged in recent studies. For example, Meslamani *et al.* recently reported an automated work flow, using several methods to optimally browse target ligand space according to existing knowledge on either ligand and target space under investigation [8].

> *"Cloud computation will have profound ability to benefit the scientific community by helping break the knowledge barrier, reducing costs, enhancing productivity and accelerating the progression of research by consolidating existing or newly developed data and algorithms/tools."*

NIH newly launched a project of "development of a knowledge management center for illuminating the druggable genome" with a goal to "increase the understanding of the properties and functions of poorly characterized and/or unannotated proteins within the most commonly drug-targeted protein families", and a project of "NIH big data to knowledge" for "enabling biomedical scientists to capitalize more fully on the big data being generated by those research communities". Such projects will certainly benefit computational target fishing.

Cloud computation is the fashion

Computational target fishing is an extraordinary computational power and store resource demanded job. For example, docking a small molecule to the x-ray structures stored in whole PDB database will take hours or days to complete the task. Given the exponential increase in computer power, advancing technologies in virtual hardware platform and powerful parallel computation on general use graphical processing units [9], the deployment of the cloud computation for computational target fishing is feasible. The cloud computation will be used to disseminate the algorithms/software and related resources that permit *in silico* target identification, side effect prediction, repurposing, polypharmacological studies of known drugs and absorption, distribution,metabolism, excretion and toxicity prediction. It can be accessible from all popular user terminals such as PC, tablet, pad or smart phone, since all the calculation is performed on the cloud computing server. Given these functions into consideration, it would be more convenient for experimental chemists or biologists, particularly for those with little knowledge regarding high-performance computing techniques or machine learning algorithms.

As such, cloud computation will have profound ability to benefit the scientific community by helping break the knowledge barrier, reducing costs, enhancing productivity and accelerating the progression of research by consolidating existing or newly developed data and algorithms/tools. Actually, such cloud computation services for target identification have emerged online. For example, DOCK Blaster makes the docking process automatically and provides public access service for structure-based ligand discovery [10].

Collaboration is the trend

Computational target fishing is naturally collaborative work as it links the cheminformatists, chemists, biologists and pharmaceutical scientists by the computational approaches, compounds, biology activity and drugs, respectively, to speed up drug discovery. In the last few years, we have witnessed more and more collaborations between academia and drug

industry. For example, big pharmaceutical companies like Glaxo-SmithKline, Genentech, Merck, Abbott, Pfizer, Novartis, AstraZeneca and Eli Lilly provided information of their proprietary compounds for drug repurposing [11]. More recently, scientists called for the release of patient-level clinical trial data to benefit the drug discovery by increasing efficiency of drug development and reducing the duplication of efforts [12]. In the meanwhile, scientists working on this field should be able to provide such facilities to guide the collaborations with the help of advance information technologies. For example, TargetHunter offers BioassayGeoMap function to help find potential collaborators nearby with bioassays established for target validation [13].

Conclusion

Advances in medicinal chemistry have produced extremely large libraries of potential therapeutic ligands. Simultaneously, advances in cell signaling, chemical genomics and proteomics have provided new potential drug targets, whereas computational target fishing technologies increase our ability to efficiently and effectively screen against these targets in high-throughput formats. With integrated data and tools, cloud computation technologies, and collaboration with scientists working on biochemistry, pharmacology, pharmacokinetics and toxicology, computational target fishing will facilitate drug discovery and drug design to address the unmet medical needs.

Financial & competing interests disclosure

The authors would like to acknowledge the financial support for the laboratory at University of Pittsburgh from the NIH DA025612 and HL109654 (XQ Xie). The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

References

1 Keiser MJ, Setola V, Irwin JJ *et al.* Predicting new molecular targets for known drugs. *Nature* 462(7270), 175–181 (2009).

2 Nidhi, Glick M, Davies JW, Jenkins JL. Prediction of biological targets for compounds using multiple-category Bayesian models trained on chemogenomics databases. *J. Chem. Inf. Model.* 46(3), 1124–1133 (2006).

3 Li H, Gao Z, Kang L *et al.* TarFisDock: a web server for identifying drug targets with docking approach. *Nucleic Acids Res.* 34(web server issue), W219–W224 (2006).

4 Cheng T, Li Q, Wang Y, Bryant SH. Identifying compound-target associations by combining bioactivity profile similarity search and public databases mining. *J. Chem. Inf. Model.* 51(9), 2440–2448 (2011).

5 Jenkins JL, Bender A, Davies JW. *In silico* target fishing: predicting biological targets from chemical structure. *Drug Discov. Today* 3(4), 413–421 (2007).

6 Rognan D. Computational approaches to target fishing and ligand profiling. *AIP Conf. Proc.* 1456, 157–164 (2012).

7 Rognan D. Structure-based approaches to target fishing and ligand profiling. *Mol. Inform.* 29(3), 176–187 (2010).

8 Meslamani J, Bhajun R, Martz F, Rognan D. Computational profiling of bioactive compounds using a target-dependent composite workflow. *J. Chem. Inf. Model.* 53(9), 2322–2333 (2013).

9 Ma C, Wang L, Xie XQ. GPU accelerated chemical similarity calculation for compound library comparison. *J. Chem. Inf. Model.* 51(7), 1521–1527 (2011).

10 Irwin JJ, Shoichet BK, Mysinger MM *et al.* Automated docking screens: a feasibility study. *J. Med. Chem.* 52(18), 5712–5720 (2009).

11 Chong CR. Repurposing drugs for tropical diseases: case studies and open-source screening initiatives. In: *Drug Repositioning: Bringing New Life to Shelved Assets and Existing Drugs.* Frail DE (Ed.). John Wiley and Sons, Inc., NJ, USA, 347–373 (2012).

12 Eichler HG, Petavy F, Pignatti F, Rasi G. Access to patient-level trial data – a boon to drug developers. *N. Engl. J. Med.* 369(17), 1577–1579 (2013).

13 Wang L, Ma C, Wipf P, Liu H, Su W, Xie X-Q. TargetHunter: an *in silico* target identification tool for predicting therapeutic potential of small organic molecules based on chemogenomic database. *AAPS J.* 15(2), 395–406 (2013).

Websites

101 ChEMBL database. www.ebi.ac.uk/chembldb

102 PubChem database. http://pubchem.ncbi.nlm.nih.gov

103 Potential drug target databases. www.dddc.ac.cn/pdtd

104 DrugBank. www.drugbank.ca

105 Therapeutic target database. http://xin.cz3.nus.edu.sg/group/ttd/ttd.asp

106 PharmGKB. www.pharmgkb.org

107 DrugMap. http://r2d2drug.org/DMC.aspx

108 PDB. www.pdb.org

109 Human protein atlas. www.proteinatlas.org

110 MetaCore. http://thomsonreuters.com/metacore

111 KEGG. www.genome.jp/kegg

112 Comparative toxicogenomics database. http://ctdbase.org

113 Target-toxin database. www.t3db.org