Bioinformatics, 31(12), 2015, 2049–2051
doi: 10.1093/bioinformatics/btv099
Advance Access Publication Date: 16 February 2015
Applications Note

Systems biology

TarPred: a web application for predicting therapeutic and side effect targets of chemical compounds

Xian Liu[1,†], Yuan Gao[2,†], Jianlong Peng[1], Yuan Xu[1], Yulan Wang[1], Nannan Zhou[3], Jing Xing[1], Xiaomin Luo[1], Hualiang Jiang[1,3,*] and Mingyue Zheng[1,*]

[1]State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China, [2]Department of Applied Mathematics, University of Washington, Seattle, WA 98195-3925, USA and [3]State Key Laboratory of Bioreactor Engineering and Shanghai Key Laboratory of Chemical Biology, School of Pharmacy, East China University of Science and Technology, Shanghai 200237, China

*To whom correspondence should be addressed.
[†]The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.
Associate Editor: Jonathan Wren

Received on January 1, 2015; revised on February 9, 2015; accepted on February 10, 2015

Abstract

Motivation: Discovering the relevant therapeutic targets for drug-like molecules, or their unintended 'off-targets' that predict adverse drug reactions, is a daunting task by experimental approaches alone. There is thus a high demand to develop computational methods capable of detecting these potential interacting targets efficiently.
Results: As biologically annotated chemical data are becoming increasingly available, it becomes feasible to explore such existing knowledge to identify potential ligand–target interactions. Here, we introduce an online implementation of a recently published computational model for target prediction, TarPred, based on a reference library containing 533 individual targets with 179 807 active ligands. TarPred accepts interactive graphical input or input in the chemical file format of SMILES. Given a query compound structure, it provides the top ranked 30 interacting targets. For each of them, TarPred not only shows the structures of three most similar ligands that are known to interact with the target but also highlights the disease indications associated with the target. This information is useful for understanding the mechanisms of action and toxicities of active compounds and can provide drug repositioning opportunities.
Availability and implementation: TarPred is available at: http://www.dddc.ac.cn/tarpred.
Contact: hljiang@simm.ac.cn or myzheng@simm.ac.cn

1. Introduction

An emerging avenue that brings system biology to drug discovery is polypharmacology, which hypothesizes that most drugs exert their effects via multi-target interactions (Hopkins, 2007). Such drug promiscuity has also led to unwanted and unexplained drug reactions. It is therefore in high demand to develop *in silico* approaches for identifying potential biological targets for chemical compounds, which can be used to: (i) obtain small molecules with optimum therapeutic effects, (ii) explore ligand–target interactions and their related biochemical mechanisms and (iii) increase research productivity towards valid drug repositioning. Recently, the rapid growth of bioassay data volume has brought new opportunities to *in silico* target prediction. A plethora of ligand-based target prediction models has been proposed (Koutsoukas *et al.*, 2011). For example, the

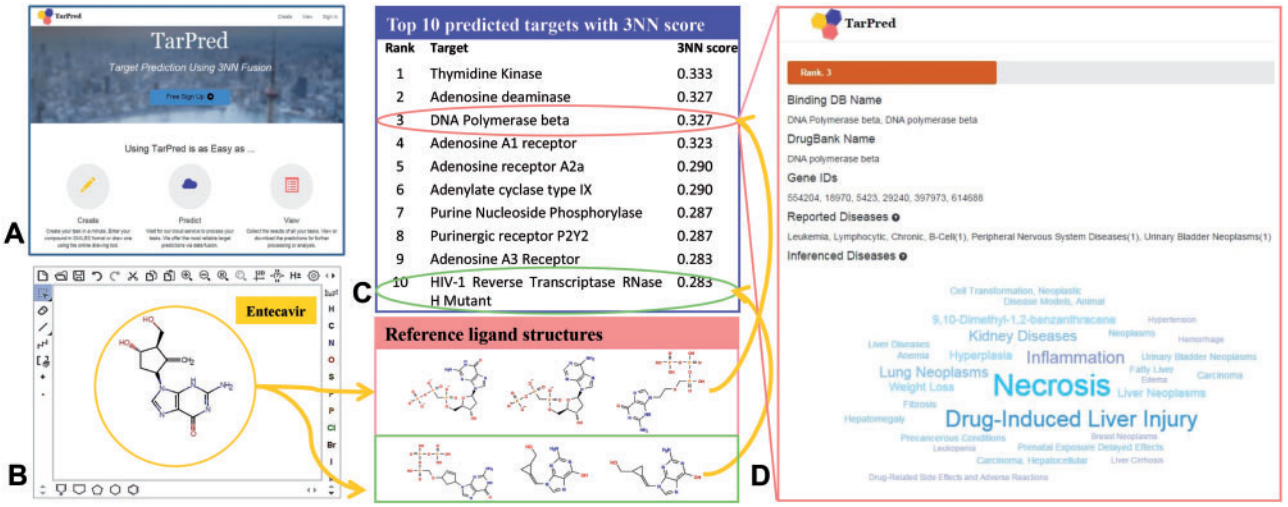

Fig. 1. (**A**) Screenshot of TarPred. (**B**) The chemical structure input interface. (**C**) Predicted target profile for Entecavir and the most similar structures for two potential targets. (**D**) Detailed target profile for DNA polymerase beta

similarity ensemble approach (SEA) (Keiser *et al.*, 2007) is a featured one among them that quantitatively estimates pharmacological relationships of targets by the set similarities of their respective ligands. These relationships have led to the discovery of many new uses and side effects for drugs. By compiling a large reference library containing 533 drug relevant target families, 179 807 individual ligands and a total of 246 053 binding data between them, we initially developed an efficient and accurate target fishing scheme using K Nearest Neighbor (KNN)-based data fusion with molecular similarity searching (Liu *et al.*, 2014). Compared with SEA, the proposed approach exhibited superior performance on predicting both therapeutic and 'off' targets for the approved drugs from DrugBank and Therapeutic Target Database (TTD). It also showed excellent results on other external validations that try to identify newly reported ligand–target interactions or to predict side effect targets for drugs withdrawn due to human Ether-à-go-go-Related Gene (hERG) toxicity. To bring more benefits of this approach, we here present a user-friendly web application that permits easy and effective drug targets prediction.

2 Approach and implementation

2.1 Data organization

In addition to target prediction, we also provide associated diseases information for each target, which can in turn be used to validate the predictions, or to discover potential novel usages of drugs. The Comparative Toxicogenomics Database (CTD) (Davis *et al.*, 2014) was used for data integration (which provides gene–disease relationships extracted from the literature), with the following steps:

1. The target families of TarPred cover 1703 individual protein sequences, which were used to perform Basic Local Alignment Search Tool (BLAST) search for their corresponding genes and Uniprot gene IDs.
2. Among the 533 target families, 359 were able to find their associated diseases with direct evidences, where a target is considered related to a disease if the corresponding genes are biomarkers of or are in the etiology of the disease.
3. In addition to direct target-disease relationships, diseases predictions for 417 targets using CTD-curated chemical-gene interactions were also provided, where each predicted disease has an inference score indicating confidence.

2.2. Inputs

Figure 1A shows a screenshot of the landing page of TarPred, which exposes the access to all major functionalities. TarPred allows users to input a compound in SMILES format or to draw one using Marvin JS (http://www.chemaxon.com/products/marvin/) (Fig. 1B). The compound is then sent to the server for 3NN fusion score calculation using JChem command line tools (http://www.chemaxon.com/products/jchem-base/). The computation takes around 7 min to complete due to the size of our library, and users can always track the process of their tasks at results page.

2.3 Outputs

After the computation is completed, users can view an interactive ranking list containing the top 30 predicted targets with 3NN scores. For each target, detailed information about its Binding DB and DrugBank names, Gene IDs, predicted diseases and the pictures of three nearest compound neighbors to the query compound in the target ligand set are available. Users can also download the results in tsv format for post processing and analysis.

3 Case study

Entecavir is a nucleoside analog drug that was used in the treatment of chronic hepatitis B virus infection by inhibiting DNA polymerase (Levine *et al.*, 2002). Recent studies revealed that Entecavir also inhibits human immunodeficiency virus (HIV)-1 infectivity at clinically relevant concentrations, as shown by the temporal association of Entecavir therapy with approximately 1-log 10 reductions in HIV-1 RNA levels *in vivo* and *in vitro* studies (McMahon *et al.*, 2007). These on-target and off-target interactions of Entecavir can be successfully captured by TarPred. As shown in Figure 1C, the DNA polymerase beta is ranked the third, and HIV-1 reverse transcriptase RNAse H is identified at the 10th ranking place. The output target profile for DNA polymerase beta is shown in Figure 1D.

Funding

This work was supported by the National Natural Science Foundation of China [grants 21210003, 81230076 and 81430084], Hi-Tech Research and Development Program of China [grant 2012AA020302] and the National Science and Technology Major Project 'Key New Drug Creation and Manufacturing Program' [grant 2014ZX09507002].

Conflict of Interest: none declared.

References

Davis,A.P. *et al.* (2014) The Comparative Toxicogenomics Database's 10th year anniversary: update 2015. *Nucleic Acids Res.*, **43**, D914–D920.

Hopkins,A.L. (2007) Network pharmacology. *Nat. Biotechnol.*, **25**, 1110–1111.

Keiser,M.J. *et al.* (2007) Relating protein pharmacology by ligand chemistry. *Nat. Biotechnol.*, **25**, 197–206.

Koutsoukas,A. *et al.* (2011) From in silico target prediction to multi-target drug design: current databases, methods and applications. *J. Proteomics*, **74**, 2554–2574.

Levine,S. *et al.* (2002) Efficacies of entecavir against lamivudine-resistant hepatitis B virus replication and recombinant polymerases in vitro. *Antimicrob. Agents Chemother.*, **46**, 2525–2532.

Liu,X. *et al.* (2014) In silico target fishing: addressing a "Big Data" problem by ligand-based similarity rankings with data fusion. *J. Cheminform.*, **6**, 33.

McMahon,M.A. *et al.* (2007) The HBV drug entecavir—effects on HIV-1 replication and resistance. *N. Engl. J. Med.*, **356**, 2614–2621.