Drug Target Identification by Maximizing Information Flow

Bingjie（Helen）Zhou

under the direction of

Dr. Shao Li

Bioinformatics Division and Center for Synthetic and Systems Biology

Tsinghua University

Research Science Initiative-Tsinghua

August 16, 2015

Abstract

Identifying drug targets is the first step and also a critical step for modern drug discovery and development of novel therapeutic approaches in pharmacology. Likewise, identifying new drug targets is crucial to drug discovery both in academic and pharmacology areas. During past decades, researchers have paid much more attention to the identification and validation of targets. However, up to now, only 324 targets have been discovered for clinical drugs [13]. It is imminent to find an efficient way to identify targets. When considering that similar proteins may have similar functions, it has been assumed new targets can be predicted from similar proteins and known drug-target interactions. In this study, we devise a drug-protein network with known drug-drug interactions, drug-target interaction and protein-protein interactions. From this network, every protein and drug could connect through connections to other drugs and proteins. Using an algorithm called maximizing information flow (MAXIF), we can then compile to calculate the association of each drug-protein pair. The MAXIF method simulates drug-target effect on the computer and has a fast speed of discovery potential targets. Leave-one- out cross validation is used and the AUC (area under ROC (receiver operating characteristic curve)) which is a standard to judge the effectiveness of a method is 0.902 which demonstrates the method has a high level of accuracy and could be applied to target identification [14]. So with the MAXIF approach, drug discovery processes targeting proteins will be greatly speed up. We hope that numerous potential applications of drugs will be discovered using this methodology and side effects and toxicity of drugs will become more understandable, allowing for substantial improvement in drug research and development.

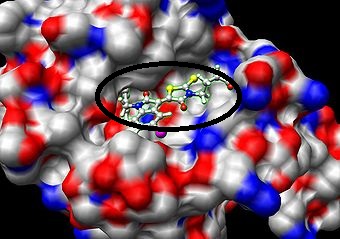
**Summary**

Drug targets play important roles in the effectiveness of drugs, but up to now, only 324 targets have been identified. The lack of knowledge of targets has been huge barrier in the development of drugs and treatments. We propose a new method to predict targets. We know there are lots of similarities between drugs and drugs, so an idea that similar drugs have similar effectiveness came up and the same for proteins. Based on this assumption, we set up a network of of drugs and proteins with lines record interactions of them. From the network, we could clearly see each interactions between drugs and drugs, between proteins and proteins and between drugs and proteins. It is a network so that we could find interactions between any pair of drug and protein through other drugs and proteins connecting to them. Then MAXIF algorithm which used to calculate any pair of interaction score in a network is applied to calculate the interaction between drug and protein we want. According the score, it could be judged whether a target or not. Through leave-one-out validation, we find that MAXIF method has great accuracy, sensitivity and specificity and could be applied to predict targets effectively. Predicting targets is useful for finding new applications of existing drugs by finding unknown targets which could be combined to the drug and side effects could also be understood more exactly because they are caused by unwanted drug-target combinations.

1. **Introduction**

An ideal drug target is a native protein in the body whose activity is modified in a specific and controllable manner resulting in a specific effect [1]. Over eighty percent of drug targets are protein and almost fifty percent belong to GPCRs, serine, threonine and protein tyrosine kinase, MMPs, serine protease, hormone nuclear receptor and PDEs. In theory, proteins that act as drug targets must be capable of combining small molecules associated to diseases with appropriate chemical characters and affinity and could express in pathology cells and tissues resulting in therapeutic effect [2]. For example, Alzheimer’s disease has complex etiology and it attacks many molecular segments, which provides a multitude of potential drug targets. Moreover, our research aims to reduce side effects and toxicities of drugs, caused by unwanted adverse effect of targets

.



**Figure 1: drug target combining** small molecule drugs usually combine with fit sites on

targets using hydrogen bonds

1.1 Significance

Though great efforts have been paid to target identification, only 324 drug targets have been identified for clinical drugs up to now [13], which indicates that currently pharmacological industry relies on a small set of targets compared to large amount of proteins. On the other hand, a large amount of drugs can be attributed to the wrong target at the early pipeline stages which may cause undesired side effects. Thus, predicting new drug targets to address novel therapies is extremely valuable in disease treatment, side effect analysis, as well as reduction of time and experiments costs in drug development.

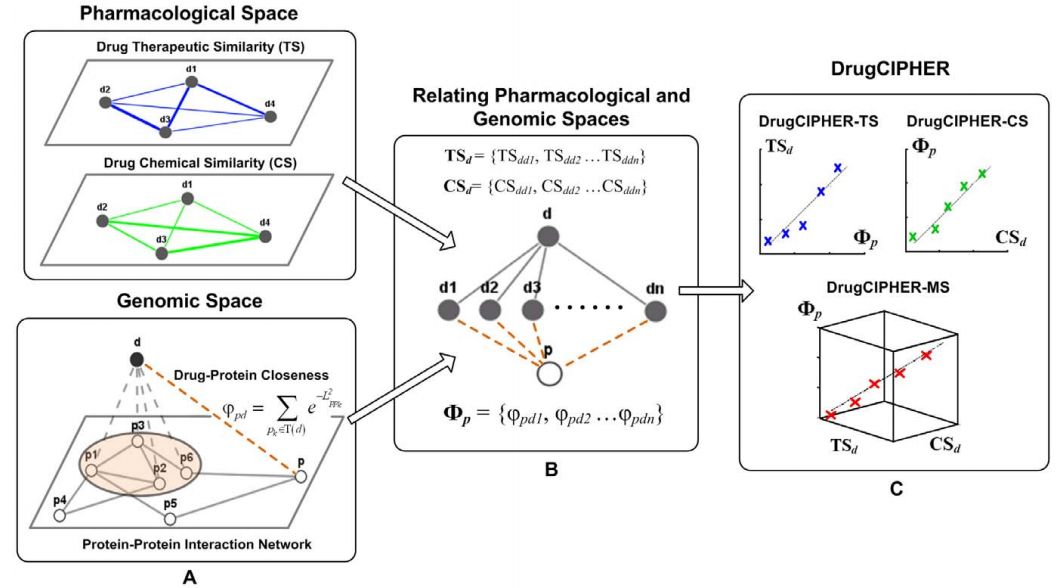
1.2 Revolution of type of drug-target interaction

In the most basic model, we assume that one drug acts selectively to one specific target. However, recent research has revealed a far more complex figure of drug reaction in the past decades. An elegant new theory by Yildirim illustrates that not only are there multiple drugs relating to one target but that one target could associate to many drugs [3]. Furthermore, this theory suggests that drugs and targets act as networks and affect each other mutually. The new theory changes effects people think of drugs and demonstrates that one drug could have lots of different effectiveness.

1.3 Methods of previous scientists

Although many effective analyses such as phenotypic effect and chemical-based approaches have previously existed, there are apparent shortcomings. For the phenotypic approach, off-target responses were activated due to similar pathways or biological processes. For the chemical approach, researchers often do not have enough prior information for major chemicals to know downstream effects. Consequently, there is urgent need to combine phenotypic and chemical indexes together, to achieve an improved method to predict drug-target association in on a larger scale [4].

Taking these observations into account, the Cipher model was proposed and validated as the most accurate way to predict drug targets, given both phenotypic and chemical indexes. .More specifically, the Cipher model assumes that similarities in pharmacological space (termed as Therapeutic Similarities (TS)) and drug chemical similarities (CS) are correlated with the relatedness of targets based on protein-protein interaction(PPI) network in genomic space. Based on this assumption, a computational framework drugCipher was built to relate pharmacological and genomic spaces with multi-dimensional information and used three linear regression models to calculate concordance scores which used as standard of correlation between drug and protein [4].



**Figure 2: Principle of drugCipher** Drugs are solid nodes and presented by ‘**d**’; proteins are hollow nodes and presented by ‘**p**’. A). Drug Therapeutic Similarity (**TS**) (blue solid edges) and Drug Chemical Similarity (**CS**) (green solid edges) comprise the pharmacological space. The protein protein interaction (**PPI**) (gray solid edges) network represents the information in the genomic space. Together with drug-target interactions (gray dashed edges), the closeness (brown dashed edges) is defined to associate a drug with any arbitrary protein. B). For drug d and protein p, two similarity vectors for d in pharmacological space (**TSd** and **CSd**) and one closeness vector for **p** (**Wp**) are constructed. C. The concordance scores between drug **d** and protein **p** are computed based on three linear regression models, which assume linear correlations exist between TSd and Wp,Wp and CSd, Wp and the combination of TSd and CSd.[4]  
 Although Cipher has a high sensitivity and precision score among all approaches, the linear regression model is not clear under the situation that the protein-protein interactions and drug-drug similarities are low, showing on the left, top and right, bottom part of graph. Likewise, the approaches above largely depended on PPI networks to estimate similarity. Since these approaches consider the shortest path as the optimal path between proteins and may overlook others paths, the reliability of the optimal path and the final result may be adversely affected [5].

1.4 MAXIF method

Motivated by these observations, we propose a new combinatorial approach to calculate score of association between a pair of drug and protein in this paper. In our approach, we first introduce a threshold to select out high similarities which have biological meanings. Then we set up a drug-protein network by integrating given drug-target interactions, PPI network information and drug-drug similarity. The magnitudes of associations are recorded under the assumption that drugs with high similarities have stronger associations as the capacity of each edge for final calculation. Then we will judge the strength of the relation based on the amount of information flow between drug and protein and we will develop the MAXIF (Maximizing information flow) algorithm in the drug-protein network system to figure out the path containing the maxima information and compute the max information flow as the concordance score of this pair of drug-protein. Proteins with high scores are considered as potential targets. The level of sensitivity and specificity of our method were showed by leave-one-out cross validation and demonstrate that our method has a high level of accuracy, so we could use it to predict potential targets which could excavate novel applications of existing drugs and narrow the range of test experiments, to pinpoint drugs with a specific curative effect. If we generate enough data for such targets, we could minimize problems of side effects and raise efficiency of drug-based therapies. Thus, identifying drug targets in a more efficient way will act as an important stepping-stone in pharmacology.

2. Materials and Methods

2.1 Construction of network

2.1.1 Data preperation

To construct the drug-protein network, we need to prepare data including drug-drug similarity, protein-protein interaction(PPI) and known drug targets interaction first.

First, 6810 drugs information were obtained including chemical structure for every drug. We assumed that drugs with more similar structures have more connections so converting extent of chemical similarity into exact score is necessary for calculating. OpenBabel[7],a chemical computing tool could convert one chemical structure format to a zero-one series and FP3[8] which defines 55 kinds of substructures and records as either one when the substructure is in the chemical structure of drug or zero when it is not in is used. By turning chemical structures into FP3 format and comparing 0-1 series, we could calculate out corresponding similarity score. We finally got 23177836 pairs of drug similarity scores.

For protein-protein interaction, protein-protein comparisons were gained from the String [10] website in July, 2015. Protein-protein pairs with confidence score higher than 700 were considered as connected protein-protein according to literature [9] and we ultimately obtained 650466 pairs of proteins.

From drug information files, known targets and experimental targets of each drug were recorded and we obtained 15305 pairs of drug-target links. The drug-target links are significantly important because they connect protein-protein networks and drug-drug networks to form the final interactions between drugs and proteins.

2.1.2 Threshold and construction of network

Since most small similarity scores for drug drug interactions are disturbing and only high similarity scores have true biological meanings, we set up a threshold α for drug-drug similarity scores which means we only use drug pairs with similarity scores higher than or equal to the threshold. [5] Drug target interactions have been demonstrated already, we must provide the influence of the known drug-target flows in drug target connections. We set a very large coefficient β which could reach millions for drug target interaction score to provide unblocking flow between known drug-target interaction. For protein-protein interaction, we set the coefficient γ to make number of interaction scores could be compared directly.

Then we regard every drug and protein as a node in the network and record every edge linking nodes. Every edge has attributes including the name of beginning drug or protein, next nodes connecting to it, capacity of the edge which is the interaction score, current flow in this arc, and reverse arc of the edge. With these definitions, the drug-protein network is denoted as indirected graph G=(V,E) in which V means vertexes and E means edges. Each edge E (u,v) (from u to v) has a positive capacity c(u,v) which record their maximum flow in the edge.

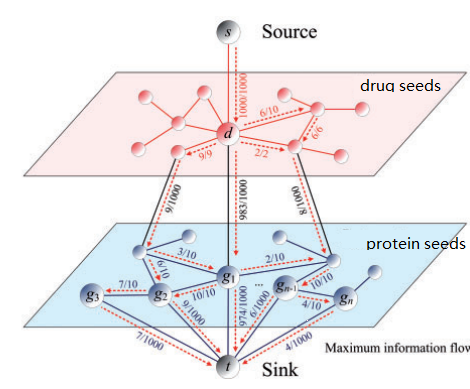


Figure 3: Model of constructed network. Nodes are proteins and drugs, edges are interaction and numbers on edges are capacities representing the extent of the connection. Two layers mean separatel protein networks and drug networks.(change from figure1 in [5])

2.2 Maximizing information flow in network to obtain interaction score

2.2.1 Explanation and properties of network flow

Our purpose is to identify drug targets with other known drug-drug, protein-protein and drug-protein information. The process is just like a substance flows from source point to the sink point by consuming energy in a network. By coincidence, model of network flow has the same principal. And we wonder the strongest association, so we use a method called Maximizing information flow(MAXIF) which is a special computation in network flow to calculate out largest association score between arbitrary pair of drug and protein using the drug protein network.

Each directed edge of the flow network could be considered as tubes to transport substances. Every substance has its fixed capacity which we define as the interaction score we first imported. Vertexes are connecting points of tubes. Two vertexes, source and sink, are special because source is the point where all flows come out and sink is where all flows converge. Except source and sink points, flows only pass these vertexes instead of accumulating. In other words, the amount of substance flowing into the node is equal to amount flowing out. The characteristic is called flow conservation [11]. Maximizing information flow is a specific problem in network flow. It’s goal is to calculate the maximum flow to transport a substance from source to sink. In our project, the goal is converted to calculate out the maximum interaction score between a pair of unknown drug and protein which is quite similar to theMAXIF model. Here are some properties in network flow:

(1) current flow of each edge is lower than its capacity

(2) reverse flow of edge is equal to its opposite number

(3) total amount of flow from source is equal to total amount of flow entering sink

2.2.2 Implement of MAXIF

The input of our MAXIF program includes total number of proteins and drugs which are considered as nodes in the network n and total number of pairs which are considered as edges m.And the drug-protein network G = ( V , E ) including begin node u[i] ,end node v[i] and capacity of edge c ( u , v ) flowing from u to v. First, we incorporate the source vertex s and define each drug a directed edge of infinite capacity pointing from source vertex s to drug vertexes d which means there is no connection between the nodes. Similar, we incorporate a sink vertex t and define each protein a directed edge of infinite capacity pointing from each protein vertex p to sink vertex t. We obtain a directed graph G = ( V , E ) , where V = V ∪ {s,t} and E = E ∪ { ( s , d ) ∪ ( p , t ) }. We also define capacity function c ( u , v ) = c ( v , u ) and c ( s , d ) = c ( p , t ) = ∞ . Then we input the network file which contains nodes and capacities and build up the drug protein network with exact capacity of every edge.

It is obvious that the network is a flow network if drug and protein networks are connected.And there is must at least one route of link between each drug and each protein. We regard G as a information flow network and interpret information flow f as a scheme of distribution of total amount of information from source s over all the edges in the network such that total amount of flow from source is equal to total amount of flow accepted by sink.

The total amount of flow from source is determined by value of information flow in network.And the flow with max value in the scheme is the route we desire because the route allowing the maximum information flow from the source has the maximum relation between the pair of drug and protein. According to literature [14],a maximum flow could be calculated out by an ISAP algorithm.

2.2.3 Applications

Once the score has been calculated out, we obtained the exact association between the drug and protein and could make a deeper analyze between the drug and protein. Because the efficient algorithm and simple construction of network, we could simply get association score from the maximum flow in the network to identify whether the protein is a target or not [5]. In this way, the efficiency of the matching drug and target is raised to a large extent and also the range of experiments that scientists need to do to seek targets is reduced too which saves lots of time and cost of experiments. Additionally, identifying targets quickly give great help to novel treatment and new applications of existing drugs. Moreover, knowing all targets of a drug exactly could make us have better understanding of side effects and toxicity and avoid them which make great sense to modern medical area.

2.3 Validation methods

We perform a leave-one-out cross validation experiment to examine the capability of our method in discovering targets that are known to be associated with a certain drug from a set of candidates. First, in validation, we take a known drug target association from each run, assume that the association is not known and prioritize the target against 99 control drug that are selected randomly from proteins. To avoid contingency, we randomly generate 1000 files like this. Second, after each validation run ,we are able to gain a list of rank proteins. We then calculate rank ratios of drugs by dividing their ranks with the number of drugs in the list. Third, we are able to generate a receiver operating characteristics (ROC) curve [12] to appraise the sensitivity and specificity of the method and further calculate the area below the curve (AUC)to validate feasibility of the method. Larger AUC values indicate higher performance of a prioritization method.

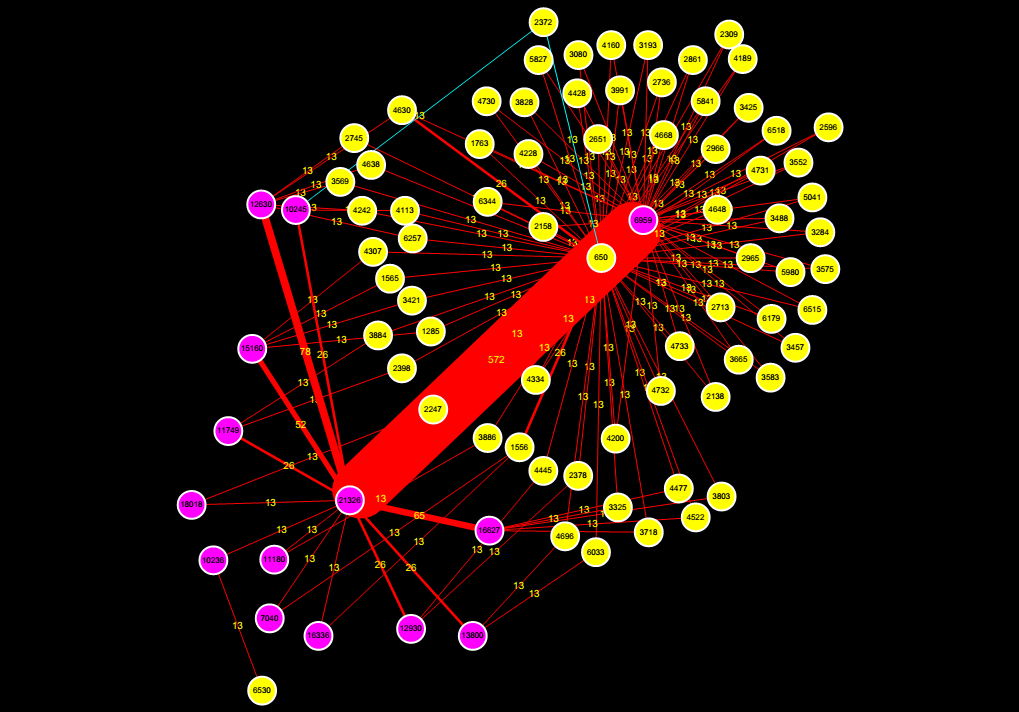


Figure 4: network generated from one of data file yellow nodes represent drugs and pink nodes represent proteins, each label on nodes is the identification number for the molecule. Lines are interactions between drugs and proteins and their widths were defined from level of association between molecules which are labeled as numbers on lines. The blue lines represent the path that MAXIF found the association between the protein and drug which pair is the only one that has been demonstrated as drug and target.

3. Results

Under the default parameters (α= 0.3, β= 10^7 , γ= 1), we obtained a drug-protein network composed of 21325 nodes (including 6810 drugs and 14515 proteins) and 38240198 edges. Using the network, we examined the performance of the MAXIF method by ranking a list of proteins including a known target.

We used leave-one-out validation experiment and generated ROC (receiver operating characteristics) curve to see the sensitivity and specificity of the method. We also calculated the area under the ROC curve (AUC) which could judge the accuracy of the method. The ROC curve is presented in Figure 5.

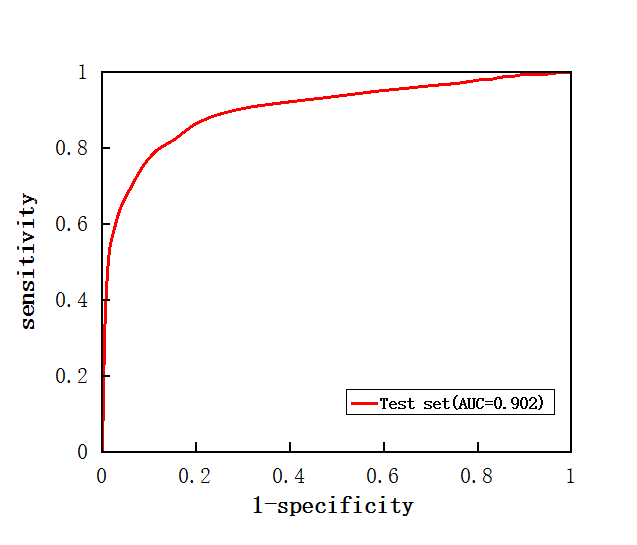


Figure 5: ROC curve for leave-one-out validation with parameters(α= 0.3，β = 10^7 ，γ = 1)

From the ROC curve, we could see that as the specificity grows up, sensitivity of method grows rapidly. Also the AUC value is 90.2% which is higher than 90% which demonstrates that MAXIF has high accuracy of prediction and could rank unknown drug-protein interactions properly [13]. From these satisfying results, we can conclude that the MAXIF method could be used to identify and predict targets and because of the high efficiency of the method, MAXIF could make great contribution to speed up discovering novel drugs and disease therapies.

4.Discussion

Single chemical indexes and therapeutic indexes previously used to see drug-protein interactions are too unilateral and could be misled by other confounding variables. Furthermore, cipher combines chemical and therapeutic indexes together, but the linear regression model could be largely affected by small interaction scores of drugs and proteins. Also the cipher approach is computationally expensive to model regression line, each time. Consequently, an accurate and more efficient method is attempted here.

The MAXIF method is based on the drug-protein network which is built by drug-drug interaction, known drug-target interaction and protein-protein interaction which comes from DrugBank and String website. In the network, we have interaction scores of molecules as weights of each edges. Then, we traverse the whole network from source point to sink point through maximum weighted edges passing through drug and protein we want to obtain exact number of interaction score. For example, when we want to find interaction score between a and b, we start from point a and find point with maximum flow to it among all next points connecting to it. We repeat the procedure and keep finding next maximum flow point continuously. If b point could not be found, we need to retrace the path and do same procedure in another path. Finally, we will find the point b and have the maximum flow path and the maximizing interaction score between a and b. The time computing 100 pairs of score is not more than 2 seconds according to real experience of my code, Thus, the efficiency of MAXIF is amazing and could speed up the research in medical area,

The key point of MAXIF in the information model is that the method based on the assumption that the relationship between a drug and protein can be captured by maximum information flow sent from source point. Although the shortest path is used most commonly to find extent of drug-protein relationship in previous methods, shortest path only considers one optimal path overlooking other paths. However, maximizing information flow method considers all paths by flowing through them from which could avoid the shortcoming of unilateral consideration of other way of connections. Moreover, the MAXIF method could also clearly provide path of maximum information which could contribute to the research of pharmacology.

Also, unlike other methods the deeply rely on one source of data, MAXIF integrates different data source like protein-protein interaction or drug similarities. In this case, it could have more complete and accurate judge of interactions instead of only depend on one kind of data.

We found that when ranking the proteins in cross validation, there are many proteins ranked similarly. Because we defined each protein-protein interaction as 1 when they are demonstrated as connecting, there is no clear differentiate of level of interaction between different pair of proteins While calculating the score passing through them, their might be no difference of different protein, so high frequency of same score appears. In further optimizations of this method, different levels of coefficients should be used and different interaction scores between proteins should be calculated through their similar of amino acid sequences or structure.

5. Conclusion

In our research, a network is built to connect all drugs and proteins directly or indirectly using all drug-drug similarities which calculated from comparison between chemical structures of drugs, known drug-target interactions and protein-protein interactions derived from DrugBank and String website. After there are connections among all drugs and proteins, we could find interaction score of any pair of drugs and proteins through other drugs and proteins. We use the Maximizing information flow method which is a computer algorithm to compute connection score of two nodes in network to calculate interaction scores of any pair of drug and protein to obtain extent of the association and to identify whether it is a target or not.

Leave-one-out cross validation method which rank an unknown drug-target pair was used to examine the accuracy and feasibility of the method and our AUC score which used to judge the accuracy of the method is 0.902 which means the method has high level of accuracy, sensitivity and specificity. Thus, We find that MAXIF could be applied to identify and predict targets and its prediction speed is much faster than other methods.

Using MAXIF in drug target identification is helpful to identify targets quickly so that more applications of existing drugs could be discovered rapidly and novel treatments of diseases could be worked out. Side effects are unwanted drug-targets interactions and after getting data of all targets of one drug, understanding of side effects and toxicity could be more profound.

6. Acknowledgements

First, I want to give my best thanks to my mentor Dr. Shao Li and co-mentor Mr.Songpeng Zu. They gave me great help no matter in the way of thinking or skills in writing codes and data analyzing. They also make me develop a very good behavior on dealing with my study files during the research. Also I want to thank my tutor Ana Lyons who gave me great help in paper writing and skills of presentation. Ana is a very serious and inspiring person.She gave me a lot of encouragements and very detail remarks of my work which may influence my writing in whole life. In addition, I want to thank RSI and CEE giving me this excellent opportunity to approach the true scientist research and make me have such a unforgettable summer. I also want to thank my friends and lab-mates in RSI. They are really creativity and friendly and truly gave me a lot of help in different area. Finally, thanks to Tsinghua university giving us such a wonderful study environment. Thank you.

References

[1] Rang HP, Dale MM, Ritter JM, Flower RJ, Henderson G (2012). ["Chapter 2: How drugs act: general principles"](http://www.inkling.com/read/rang-dale-pharmacology-ritter-flower-henderson-7th/chapter-2/protein-targets-for-drug-binding). *Rang and Dale's Pharmacology*. Edinburgh; New York: Elsevier/Churchill Livingstone. pp. 6–19. [ISBN](https://en.wikipedia.org/wiki/International_Standard_Book_Number" \o "International Standard Book Number) [0-7020-3471-1](https://en.wikipedia.org/wiki/Special:BookSources/0-7020-3471-1" \o "Special:BookSources/0-7020-3471-1).

[2]Jianxiong Yang,Zhihui Liu,Effect of drug targets in developing novel drugs,*Lishizhe Medicine and Materia Medica Research* ,vol. 20 no.3(2009)

[3]Andrew L. Hopkins,Network biology illustrates our understanding of drug actions, *Nature Biotechnology,*vol 25,no.10 (October 2007)

[4]Shiwen Zhao, Shao Li,Network-based relating pharmacological and genomic spaces for drug target identification.*Plos one ,*vol 5,issue 7(July 2010)

[5]Yong Chen,Tao Jiang,Shao Li,Uncovering disease genes by maximizing information flow in the phenome-interactome network,*Bioinformatics,* vol 27, page i167-i176(2011)

[6]Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, et al. (2008) DrugBank:  
a knowledge base for drugs, drug actions and drug targets. Nucleic Acids Res 36:  
D901–D906.

[7]O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. (2011). ["Open Babel: An open chemical toolbox"](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3198950). *Journal of Cheminformatics*

[8]unknown;”*openbabel instruct” ;Available at: http://emuch.net/html/201207/4695885.html (Accessed :July 5,2012)*

[9]Peer Bork,Lars Juhl Jensen,Christian von Mering ,”Chapter 4 :FAQ” ,String documentation(2000-2011)

[10] Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, Doerks T, Julien P, Roth A, Simonovic M, Bork P, von Mering C (2009). ["STRING 8--a global view on proteins and their functional interactions in 630 organisms"](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2686466). *Nucleic Acids Res* **37** (Database issue): D412–6.

[11]Thomas H. Cormen, Charles E.Leiserson,Ronald L.Rivest, Clifford Stein [Chapter 26:network flow] *Introduction to algorithms* (Second Edition)(ISBN :0-262-03293-7)

[12]Kai Ma (2010).”Accuracy evaluation of ROC curve in diagnosis of urinary tract infection”

# [13]****Qingliang Li, Luhua Lai”Prediction of potential drug targets based on simple sequence properties”****BMC Bioinformatics( 2007):doi:10.1186/1471-2105-8-353

[14][T Fawcett](http://xueshu.baidu.com/s?wd=author:(Tom%20Fawcett)%20&tn=SE_baiduxueshu_c1gjeupa&ie=utf-8&sc_f_para=sc_hilight=person" \t "http://xueshu.baidu.com/_blank),”An introduction to ROC analysis” [《Pattern Recognition Letters》](http://xueshu.baidu.com/s?wd=journaluri:(2403bfd1f9cc5c23)%20%E3%80%8APattern%20Recognition%20Letters%E3%80%8B&tn=SE_baiduxueshu_c1gjeupa&ie=utf-8&sc_f_para=sc_hilight=publish&sort=sc_cited" \t "http://xueshu.baidu.com/_blank" \o "《Pattern Recognition Letters》)(2006) 27(8):861–874

Appendix 1: Data source

|  |  |  |  |
| --- | --- | --- | --- |
| Type of interaction | Drug-drug  similarity | Drug-target interaction | Protein-protein interaction |
| Data source | [www.drugbank.ca](http://www.drugbank.ca) | [www.string-db.org](http://www.string-db.org) | [www.drugbank.ca](http://www.drugbank.ca) |
| Website version | Version 4.3 | Version 10 | Version 4.3 |

We obtained drug detail information from drugBank website version 4.3 which contains  7759 drug entries including 1602 FDA-approved small molecule drugs, 161 FDA approved biotech (protein/peptide) drugs, 89 nutraceuticals and over 6000 experimental drugs.and 4300 non-redundant protein

Protein-protein interaction conditions are obtained from String -dh.org website which is a database of known and predicted protein interactions and the interactions include direct (physical) and indirect (functional) associations

Appendix 2: First 30 line of Drug and protein labels

DB00014 1

DB00035 2

DB00050 3

DB00091 4

DB00093 5

DB00104 6

DB00115 7

DB00116 8

DB00117 9

DB00118 10

DB00119 11

DB00120 12

DB00121 13

DB00122 14

DB00123 15

DB00125 16

DB00126 17

DB00127 18

DB00128 19

DB00129 20

DB00130 21

DB00131 22

DB00132 23

DB00133 24

DB00134 25

DB00135 26

DB00136 27

DB00137 28

DB00138 29

DB00139 30

Appendix 3:First 50 lines of network file

36240198 21325

1 2 54

2 1 54

1 3 83

3 1 83

1 4 46

4 1 46

1 5 38

5 1 38

1 6 57

6 1 57

1 7 71

7 1 71

1 8 50

8 1 50

1 9 54

9 1 54

1 10 50

10 1 50

1 11 36

11 1 36

1 12 54

12 1 54

1 13 50

13 1 50

1 14 36

14 1 36

1 15 38

15 1 38

1 16 38

16 1 38

1 17 60

17 1 60

1 19 38

19 1 38

1 20 38

20 1 38

1 21 38

21 1 38

1 22 57

22 1 57

1 23 38

23 1 38

1 24 46

24 1 46

1 25 36

25 1 36

1 26 62

26 1 62

1 27 36

Appendix 4: Code of MAXIF algorithm

