DTscreen v1: A novel drug targets identification tool for pathogenic diseases through protein signature-based approach

Saowalak Kalapanulak[a,b], **Arporn Juntrapirom**[a], **Treenut Saithong**[a,b]

[a]*Bioinformatics and Systems Biology Program, King Mongkut's University of Technology Thonburi (Bang Khun Thian), 49 Soi Thian Thale 25, Bang Khun Thian Chai Thale Road, Tha Kham, Bang Khun Thian, Bangkok, Thailand, 10150*
[b]*School of Bioresources and Technology, King Mongkut's University of Technology Thonburi, (Bang Khun Thian), 49 Soi Thian Thale 25, Bang Khun Thian Chai Thale Road, Tha Kham, Bang Khun Thian, Bangkok, Thailand, 10150*

Abstract

Developing a new drug takes a very long period of time and cost is also extremely high. Drug development process involves many stages; however, about 30%-40% of experimental drugs fail to be as drugs in the market because of choosing an inappropriate biological target at the early stage of the development pipeline. Therefore, it is a big challenge to innovate a systematic method and bioinformatics tool for accurately identifying drug targets in order to avoid any failures in the initial stage of anti-microbial drug discovery program. In this work, we implement a published method for drug targets identification in mycobacterium metabolism, proposed by Kalapanulak in 2009, to be a user friendly and standalone tool, DTscreen v1. We aim to facilitate biologists to do the automatic screening for all possible attractive drug targets in any interested pathogen genomes. The tool is developed on Visual Basic language. In the first version, *Salmonella- enterica serovar Typhi CT18 (S. typhi)*, a causative agent of typhoid fever in human beings, is used as a case study in the DTscreen v1. For the case study, the whole genome of *S. typhi* was investigated and 115 genes were proposed as new drug targets based on: (1) their unique protein signatures when comparing with human host and (2) their essentiality reported from a large-scale experiment.

Keywords: Drug targets, Protein signature, Bioinformatics tool

Introduction and Objective

Several bioinformatics approaches have been developed for identifying drug targets against any pathogenic diseases. Behind these approaches, usual criteria for predicting drug targets are essentiality and non-homolog with human host based on sequence similarity [1,2]. Even though, non-homologous proteins are identified as drug target, they still can cause some side effects in humans because of the similarity in term of protein signatures. In 2009, a novel bioinformatics method for drug targets identification of *Mycobacterium tuberculosis* metabolism based on essential gene and protein signature information was proposed by Kalapanulak [3]. This approach has a high accuracy for predicting drug targets against the pathogen.

Therefore, in this work we implement a published method for drug targets identification, proposed by Kalapanulak in 2009, to be a user friendly and standalone tool, DTscreen v1. We aim to facilitate biologists to do the automatic screening for all possible attractive drug targets in any investigated pathogen genomes before doing their wet experiments. We hope that our developed tool will decrease discovery time and cost in the drug development process.

Materials and Methods

The DTscreen v1 was written in 'Visual Basic' language. This tool can identify the high potential drug targets based on two criteria: (1) no protein signature matching with any human protein signatures and (2) gene essentiality from large-scale wet experiments. The algorithm of DTscreen v1 is illustrated as follows.

2.1 Non-homolog finding

The first module in the DTscreen, non-homolog finding, is the main process for identifying drug targets. Comparing protein signatures corresponding to all genes of the pathogen with all human genes in the genome via InterPro accession numbers from InterPro database is the approach for receiving the preliminary proposed drug targets that are non-

homolog with human genes in term of protein signatures in order to avoid any adverse effects.

2.2 Essential genes screening

The second module is applied for the identification of drug targets by comparing the preliminary proposed drug target from the first module with the list of essential genes required by the pathogen from the large-scale wet experiments.

DTscreen and a case study

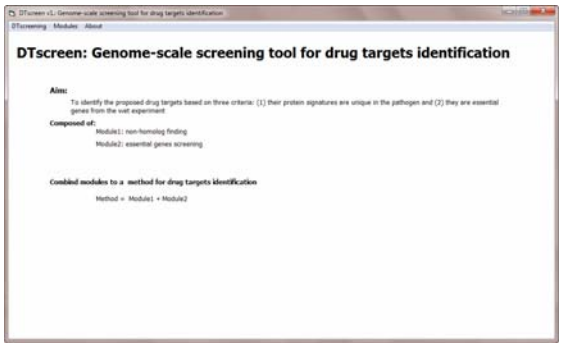

Figure 1. The main page of DTscreen v1

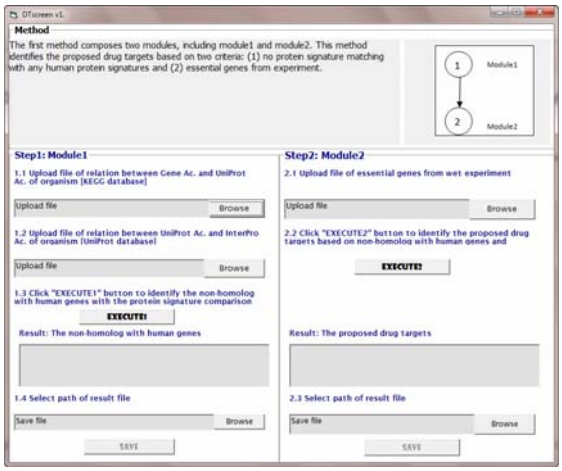

Figure 2. The method page of DTscreen v1

DTscreen v1 was tested on *Salmonella enterica serovar Typhi CT18 (S. typhi)*. The input files containing a list of genes and UniProt accession numbers and a list of UniProt accession numbers and InterPro accession numbers were downloaded from KEGG and UniProt database, respectively. Additionally, a list of essential genes from large-scale wet experiments was extracted from the literature [4]. The protein signatures of all human proteins were stored in database text file of DTscreen. The characteristics of the retrieved data and the numbers of proposed drug targets are shown in Table1 and 2, respectively.

Table 1. Characteristics of the retrieved data of S.typhi CT18 and human

Organisms	# genes in the genome	# genes with UniProt Ac.	# genes with InterPro Ac.
S. typhi CT18	4,679	4,679	3,938 (84.16%)
H. Sapiens	22,339	22,339	16,915 (75.72%)

Table 2. The number of the proposed drug targets

Organisms	# preliminary drug targets	# proposed drug targets
S. typhi CT18	2,192	115

In the case study of *Salmonella- enterica serovar Typhi CT18 (S. typhi)*, the tool can propose 115 drug targets based on non-homolog with human and gene essentiality. Interestingly, 7 of 44 current drug targets are in our 115 proposed drug targets [5].

Conclusions

- DTscreen v1 is a bioinformatics tool for identifying the whole attractive drug targets against pathogenic diseases.
- The proposed drug targets were identified based on non-homolog with human and gene essentiality.
- Users can apply DTscreen to identify drug targets in other investigated pathogens.

References

1. Singh, N.K., Selvam, S.M., Chakravarthy, P. T-iDT:Tool for identification of drug targets in bacteria and validation by *Mycobacterium tuberculosis*. In Silico biology. 2006; 6:485-493.
2. Doyle, M.A., *et al*. Drug target prediction and prioritization: using orthology to predict essentiality in parasite genomes. BMC Genomics. 2010; 11.
3. Kalapanulak, S. High Quality Genome-Scale Metabolic Network Reconstruction of *Mycobacterium tuberculosis* and Comparison with Human Metabolic Network: Application for Drug Targets Identification. Doctor of Philosophy thesis, The University of Edinburgh. 2009
4. Langridge, G.C., *et al*. Simultaneous assay of every *Salmonella Typhi gene using one* million transposon mutants. Genome Research. 2010; 19:2308-2316.
5. Becker, D., *et al*. Robust Salmonella metabolism limits possibilities for new antimicrobials. Nature. 2006; 440:303-307.