Genome-Based Screening for Drug Targets Identification: Application to Typhoid Fever

Arporn Juntrapirom[1], Saowalak Kalapanulak[1,2], and Treenut Saithong[1,2]

[1] Bioinformatics and Systems Biology Program
[2] School of Bioresources and Technology,
King Mongkut's University of Technology Thonburi, Bangkok, Thailand
arj_5104@hotmail.com, saowalak.kal@kmutt.ac.th,
treenut.sai@kmutt.ac.th

Abstract. *Salmonella enterica serovar Typhi CT18* (*S. Typhi*) is the causative agent of typhoid fever in human beings. Currently, most of the drugs used to treat this sickness have adverse side-effects. Moreover, drug-resistant strains are emerging as a serious threat for the disease. Therefore, the most effective drug targets are urgently demanded for the development of new faster-acting antibacterial agents. In this paper, a published method for drug targets identification in *Mycobacterium tuberculosis* metabolism by Kalapanulak was applied to typhoid fever. The whole genome of *S. Typhi* was investigated and 282 genes were proposed as new drug targets. Interestingly, 34 drug–affected and essential genes from the three current antibiotics are all found in our proposed drug targets.

Keywords: *Salmonella Typhi,* typhoid fever, pathogenic bacteria, genome-scale, drug targets.

1 Introduction

Typhoid fever, an infectious disease, can be called by various names, such as gastric fever, abdominal typhus, infantile remittant fever, slow fever, nervous fever, pythogenic fever, etc. In 1829, Louis gave the name of "typhoid" as a derivative from typhus [1]. Typhoid fever remains a common disease in the developing world, where it affects about 12.5 million people each year. Around 10% of them will develop severe or complicated disease. Annually, more than 600,000 people die from the disease [2] .

Salmonella enterica serovar Typhi CT18 (S. Typhi), a gram negative bacterium, causes the vast majority of typhoid fever, systemic infection, in humans. It is transmitted from a patient to a normal person by the ingestion of food or water contaminated with the feces of an infected person. The pathogen usually attacks the surface of the intestine in humans but it can develop and adapt to grow into the deeper tissues of the spleen, liver, and the bone marrow. The most common symptoms characterized by the disease often include a sudden onset of a high fever, a headache, and nausea [3]. Some patients who are infected with *S. Typhi* become life-long carriers that serve as the reservoir for the pathogen. Moreover, the causative agent has an endotoxin (which is typical of gram negative organisms), as well as the Vi antigen, which increases its

J.H. Chan, Y.-S. Ong, and S.-B. Cho (Eds.): CSBio 2010, CCIS 115, pp. 217–225, 2010.

virulence. More importantly, it is a strong pathogen for humans due to its resistance to the innate immune response system [4].

At present, patients are treated with some antibiotics that kill the *Salmonella* bacteria. Before using antibiotics, the fatality rate was 20% and deaths occurred from overwhelming infection, pneumonia, intestinal bleeding, or intestinal perforation. With antibiotics and supportive care, mortality was reduced to 1-2%. Because of appropriate antibiotic therapy, patients are usually better within one to two days and recovery within seven to ten days. Several antibiotics are used to cure the disease, for example chloramphenicol, the original drug of choice for many years. Unfortunately, because of its serious side effects, chloramphenicol has been replaced by other antibiotics. Currently, the patients have to take multiple drugs i.e., chloramphenicol, ciprofloxacin, ceftriaxone, cefexime for the treatment but most of them still have adverse effects. Moreover, drug-resistant strains emerge existing serious problem. Besides medicine, a vaccine was developed during World War II by Ralph Walter Graystone Wyckoff for prevention [5]. However it is no longer recommended for use, it has a high rate of side effects (mainly pain and inflammation at the site of the injection). Therefore, the new effective drug targets are urgently demanded for developing new faster-acting antibacterial agents.

Nowadays, biological information of *S. Typhi* is available in various resources, such as Kyoto Encyclopedia of Genes and Genomes (KEGG) database [6] collecting the gene functions of all annotated genes in any living organisms such as biochemical reactions of enzymatic genes, transport reactions of transporter genes. Additionally, Universal Protein Resource (UniProt) [7], and InterPro databases [8] are useful data resources for obtaining protein and protein signature information of living organisms, including *S. Typhi*. For InterPro database, protein signatures describing the same family or domain in terms of sequence position and protein coverage from other 10 protein signature databases are integrated into single InterPro entries. It brings us the occasion to analyze protein signatures of individual organism systematically.

Thanks to the availability of useful biological information mentioned above, it provides us an opportunity to investigate *S. Typhi* as a whole system by using systems biology approach and apply for drug targets identification. In 2009, Barh *et al.* applied comparative genomics approach for prediction of essential genes in *Neisseria gonorrhoeae* [9]. Like previous work, Dutta *et al*. applied the same *in silico* strategies for identifying drug targets of *Helicobacter pylori* [10]. Furthermore, Doyle *et al.* applied orthology-based approach for predicting essential genes in the pathogenic nematodes [11]. Additionally, in 2009 Kalapanulak proposed a novel method for identifying drug targets against *Mycobacterium tuberculosis* (*Mtb*), the causative agent of tuberculosis in humans [12]. Not only was genomic data integrated, but the wet experimental results (transposon site hybridization) were also included as a raw data for identifying drug targets in *Mtb* metabolism. Interestingly, 13 *Mtb* metabolic genes from all 42 proposed drug targets were found in a list of 70 current validated drug targets. The *in silico* approaches for identifying drugs targets have much benefit for biologists to screen for possible drug targets before doing their single-gene knockout experiment, one of the conventional methods for identifying drug targets.

In this work, a published novel method for drug targets identification by Kalapanulak in 2009 was applied to typhoid fever by integrating biological information of *S. Typhi* from various resources such as protein signatures from Interpro database, and

genome information from KEGG database. We further compared protein signatures between human host and the pathogenic organism. The proposed drug targets have been identified based on two criteria. First, their protein signatures do not match with any human protein signatures in order to reduce side effects. Second, they are essential genes required by *S. Typhi* for optimal growth from TraDis (transposon directed insertion-site sequencing) experiment.

2 Methodology

A published approach for identifying drug targets in *Mycobacterium tuberculosis* metabolism has been applied to *S. Typhi*. The whole genome of *S. Typhi* has been investigated by comparing its all protein signatures corresponding to genes in the genome between *S. Typhi* and human through InterPro accession numbers (InterPro Ac.) from InterPro database. Not only was the metabolic network of *S. Typhi* investigated, but we also did the analysis for the whole genome. Every gene in the *S. Typhi* genome has been considered as possible drug targets. The methodology for drug targets identification in *S. Typhi* is illustrated in Fig. 1.

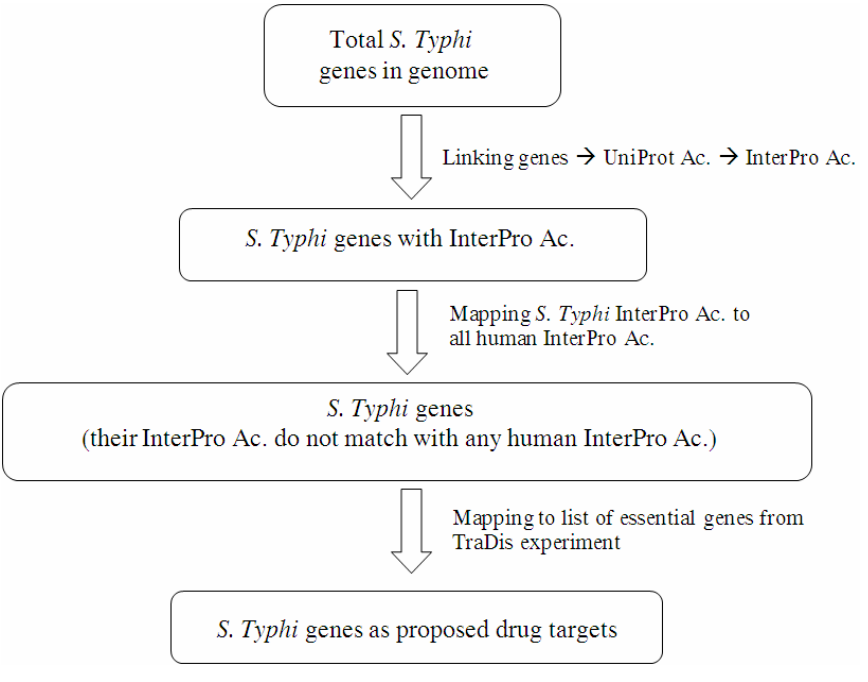

Fig. 1. The methodology for drug targets identification in *S. Typhi*

In the first step, linking genes to their protein signatures in terms of InterPro accession numbers, all genes and their UniProt accession numbers (UniProt Ac.) belonging to human (*H. Sapiens*) and *S. Typhi* genome were retrieved from KEGG database.

Next, the relations between UniProt accession numbers and InterPro accession numbers were extracted from UniProt database. Finally, the relations between genes and their InterPro accession numbers were created through UniProt accession numbers since we cannot link genes to their InterPro accession numbers directly. Hence, all genes with InterPro accession numbers were obtained and prompted for doing the protein signatures comparison between *S. Typhi* and human.

In the second step, a comparison of InterPro accession numbers between human and *S. Typhi* was made. A Visual Basic code was written for doing systematic comparison between the two groups of InterPro accession numbers. The unique InterPro accession numbers of each *S. Typhi* gene were compared with the whole set of human InterPro accession numbers. Eventually, the *S. Typhi* genes, of which the number of unmatched InterPro accession numbers are the same as the number of all InterPro accession numbers, are proposed as preliminary proposed drug targets.

In the last step, the preliminary proposed drug targets were mapped with the list of 356 essential genes required by *S. Typhi* for optimal growth from TraDis experiment (transposon directed insertion-site sequencing). Consequently, the drug targets were proposed based on two criteria: A) their protein signatures are unique in *S. Typhi*; B) they are essential genes reported from the TraDis experiment [13].

3 Results

3.1 Linking *S. Typhi* and Human Genes to Their Protein Signatures in Term of InterPro Accession Numbers

All gene names and their UniProt accession numbers of human and *S. Typhi* were downloaded from KEGG database by using the following URLs:

- ftp://ftp.genome/pub/keg/genes/organism/hsa/hsa_uniprot.list
- ftp://ftp.genome/pub/keg/genes/organism/sty/sty_uniprot.list

The results show that all genes in both of the genomes obtain UniProt accession numbers as illustrated in Table 1.

Table 1. Characteristics of Gene-UniProt-InterPro accession number relationship for *S. Typhi* and human

Organisms	Numbers of genes in genomes	Numbers of genes with UniProt Ac.	Numbers of 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%)

The relationship between UniProt accession numbers and InterPro accession numbers of *S. Typhi* and human was extracted from the UniProt database. All data were downloaded by specifying the URLs because directly downloading from their web

interface resulted in inconsistent data. The URLs, http://www.uniprot.org/uniprot?query=organism:taxID(220341)+AND+database:interpro&format=tab&compress=yes&columns=id,database(interpro) and http://www.uniprot.org/uniprot?query=organism:taxID(9606)+AND+database:interpro&format=tab&compress=yes&columns=id,database(interpro), were used to download the data of *S. Typhi* and human, respectively. Consequently, we obtained 4,679 UniProt entries for *S. Typhi*. However, 3,938 of them provided at least one InterPro accession number per entry. On the other hand, 22,339 UniProt entries were retrieved for human but only 16,915 of them have InterPro accession numbers. It should be mentioned that not all UniProt entries in both *S. Typhi* and human have InterPro accession numbers. This is because the InterPro database covers only around 84.9% of UniProtKB, for the latest public release version, Release 2010_08, in 2010 [7]. Moreover, one protein may have one or more InterPro accession numbers depending on the number of identified protein signatures. Finally, 3,938 genes of *S. Typhi* had 10,666 unique InterPro accession numbers and 72,631 unique InterPro accession numbers were obtained from 16,915 human genes.

3.2 Comparing Protein Signatures between *S. Typhi* and *H. Sapiens* via InterPro accession Numbers

The InterPro accession numbers of 3,938 *S. Typhi* genes in Table 1 were compared to 72,613 InterPro accession numbers of 16,915 human genes. We wrote the Visual Basic code for doing the systematic comparison. Eventually, we found 2,419 *S. Typhi* genes, of which all InterPro accession numbers do not match any human InterPro accession numbers, from the total of 3,938 genes (Fig. 2). These 2,419 genes are proposed as preliminary drug targets against *S. Typhi.*

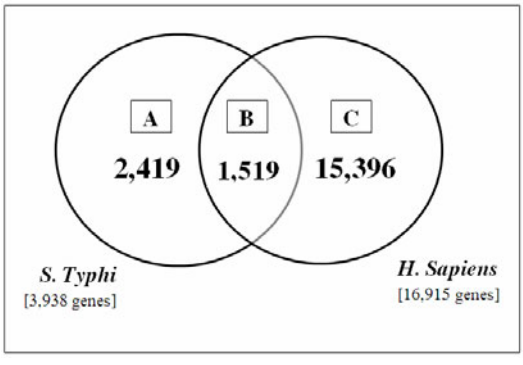

Fig. 2. Protein signature comparison between *S. Typhi* and human (*H. Sapiens*) via InterPro accession numbers; A) number of proposed preliminary drug targets against *S. Typhi*, B) number of genes that their protein signatures match between *S. Typhi* and *H. Sapiens* and C) number of *H. Sapiens* genes that their InterPro accession numbers do not match with any InterPro accession numbers of *S. Typhi*.

3.3 Mapping the List of Preliminary Drug Targets to Essential Genes from TraDis Experiment

The 2,419 genes in Fig. 2 were further compared with the list of 356 essential genes and 4,162 non-essential genes reported by Langridge *et al.* using transposon directed insertion-site sequencing (TraDis) experiment [13]. Consequently, the 2,419 genes that have been proposed as preliminary drug targets against *S. Typhi* were classified into two groups: A) 282 genes are essential, B) 2,137 genes are non-essential as shown in Fig. 3. Finally, all 282 genes in region A have been proposed as new drug targets. We further classified all proposed drug targets into seven categories based on their protein functions as shown in Fig. 4. Half of them are metabolic genes, more attractive as drug targets because of their potential for assayability and good druggability precedents. However, further analysis such as 3D-structure of proteins is still necessary for prioritizing the proposed drug targets.

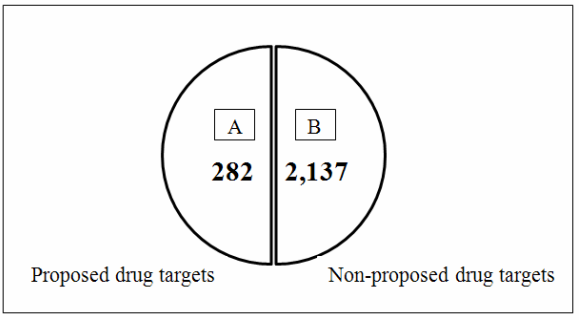

Fig. 3. Comparison between the preliminary drug targets and essential genes from TraDis experiment; A) number of proposed drug targets and B) number of non-proposed drug targets

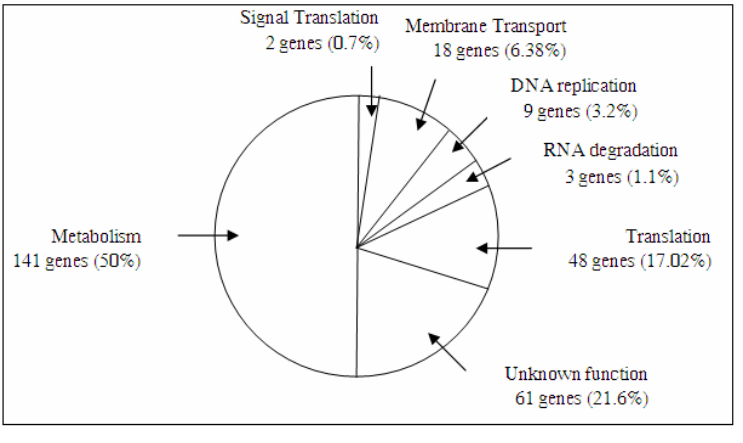

Fig. 4. Classification of 282 proposed drug targets into seven categories including Metabolism, Signal Transduction, Membrane Transport, DNA replication, RNA degradation, Translation, and Unknown function.

4 Discussion

The 282 genes from the whole genome of *S. Typhi* have been proposed as new drug targets based on two criteria: A) all of their protein signatures are different from human protein signatures; B) they are essential genes from TraDis experiment. In order to state the confidence of the novel approach for genome-scale identification of drug targets, we compared the proposed drug targets with 44 drug-affected genes reported by Becker *et al*. in 2006 [14]. The results show that 34 drug-affected genes have been found in our proposed drug targets.

In term of statistics for clarification of the confidence of these 282 proposed drug targets from the total of 3,938 investigated genes, we compared the proposed and non-proposed drug targets against 43 drug-affected genes from literature. One drug-affected gene reported by Becker *et al*. has not been included for statistical calculation because it did not have any protein signatures. Therefore, it is not include in the 3,938 investigated genes. Of the 282 proposed drug targets, 34 (12%) were current drug-affected genes. Of the 3,656 non-proposed drug targets, 9 (0.3%) were current drug-affected genes. To demonstrate the probability for obtaining 34 or more drug-affected genes from randomly selecting 282 *S. Typhi* genes without replacement from the total of 3,938 investigated *S. Typhi* genes, we calculated the hypergeometric probability for obtaining 34 or greater drug-affected genes using the below formulas [15].

$$p(x \geq 34) \quad = \quad 1 - p(x \leq 33) \tag{1}$$

$$p(x \leq 33) \; = \; \sum_{i=0}^{33} \frac{\binom{3,895}{249+i}\binom{43}{33-i}}{\binom{3,895}{282}} \; = \; 0.999999999998596 \tag{2}$$

With a population size = 3,938, sample size = 282, number of successes in the population = 43, number of successes in the sample (x) = 33.

From the probability calculation we found that the probability for receiving 34 or more drug-affected genes from randomly selecting 282 genes of 3,938 investigated *S. Typhi* genes is closed to 1×10^{-12}. It means that it is uncommon to receive 34 or more drug-affected genes when randomly sampling 282 genes from the total of 3,938 *S. Typhi* genes. Therefore, the identification of 34 drug-affected genes from 282 proposed drug targets (12%) is significant enough to confirm the quality of the published novel method in term of genome-scale analysis.

Moreover, we further compared 44 drug-affected genes with essential genes based on TraDis experiment. Interestingly, all 34 drug-affected and essential genes are all found in our proposed drug targets (Fig. 5). Among 34 drug-affected genes, 32 of them function as metabolic genes, whereas the other two are unknown genes. Therefore, we may conclude that the published novel method has high accuracy for predicting drug-affected genes that are essential for the pathogen and most of them are metabolic genes.

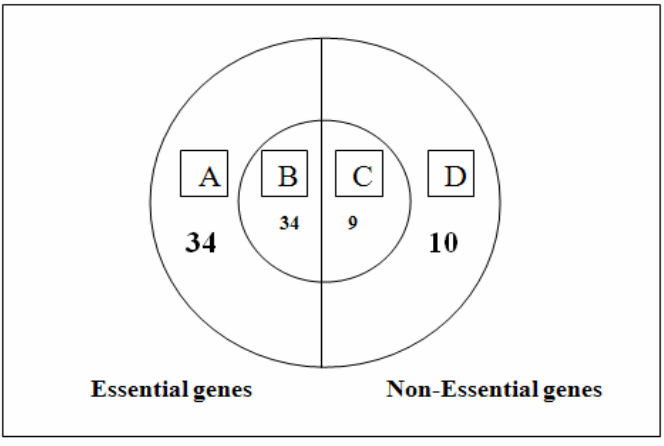

Fig. 5. Classification of 44 drug-affected genes on essential genes and non-essential genes from TraDis experiment. A) number of drug-affected genes from three current antibiotics (Becker *et al*., 2006) and essential genes based on TraDis experiment, B) number of proposed drug targets, all of them found in drug-affected and they are essential genes based on TraDis experiment, C) number of non-proposed drug targets, all of them found in drug-affected genes and they are non-essential genes based on TraDis experiment, D) number of drug-affected genes from three current antibiotics (Becker *et al*., 2006) and non-essential genes based on TraDis experiment.

5 Conclusions and Future Work

Two hundred and eighty-two genes of *S. Typhi* have been proposed as novel drug targets based on the *in silico* screening approach. Drug targets are proposed based on two criteria: A) no protein signature matching with any human protein signatures and B) essential genes from TraDis experiment. Interestingly, 34 drug–affected and essential genes from the three current antibiotics are all found in our proposed drug targets. It brings to the achievement of applying the published novel method for identifying new drug targets against typhoid fever. For future work, we will further improve the published method in order to overcome the limitation of the method. Genes without protein signature information were not included in the analysis, even though they can be interesting drug targets. Moreover, the modified method will be implemented for drug targets identification as a web application tool. It will facilitate biologists to do computational screening via our user-friendly tool before doing their wet experiments.

References

1. Kaye, K.S., Kaye, D.: Salmonella infections (including typhoid fever). In: Goldman, L., Ansiello, D. (eds.) Cecil Medicine, 23th edn. Saunders Elsevier, Philadelphia (2007)
2. Ivanoff, B., Levine, M.M.: Typhoid fever: continuing challenges from a resilient bacterial foe. Bulletin de l'Institut Pasteur 95, 129–142 (1997)

3. Shanahan, P.M., Jesudason, M.V., Thomson, C.J., Amyes, S.G.: Molecular analysis of and identification of antibiotic resistance genes in clinical isolates of *Salmonella typhi* from India. J. Clin. Microbiol. 36, 1595–1600 (1998)
4. Everest, P., et al.: The molecular mechanisms of severe typhoid fever. Trends in Microbiology 9, 316–320 (2001)
5. International Union of CRYSTALLOGRAPHY, http://ww1.iucr.org/people/wyckoff.htm
6. Ogata, H., et al.: KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 27, 29–34 (1999)
7. Apweiler, R., et al.: The Universal Protein Resource (UniProt) in 2010. Nucleic Acids Res. 38, D142–D148 (2010)
8. Hunter, S., et al.: InterPro: the integrative protein signature database. Nucleic Acids Res. 37, D211–D215 (2009)
9. Barh, D., Kumar, A.: In silico identification of candidate drug and vaccine targets from various pathways in Neisseria gonorrhoeae. Silico. Biol. 9, 225–231 (2009)
10. Dutta, A., et al.: In silico identification of potential therapeutic targets in the human pathogen *Helicobacter pylori*. In Silico. Biol. 6, 43–47 (2006)
11. Doyle, M.A., et al.: Drug target prediction and prioritization: using orthology to predict essentiality in parasite genomes. BMC Genomics 11, 1–14 (2010)
12. Kalapanulak, S.: High Quality Genome-Scale Metabolic Network Reconstruction of Mycobacterium tuberculosis and Comparison with Human Metabolic Network: Application for Drug Targets Identification. PhD Thesis. School of Informatics, The University of Edinburgh (2009)
13. Langridge, G.C., et al.: Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants. Genome Res. 19, 2308–2316 (2009)
14. Becker, D., et al.: Robust *Salmonella* metabolism limits possibilities for new antimicrobials. Nature 440, 303–307 (2006)
15. StatTrek Teach yourself statistics, http://www.StatTrek.com