NOVEL TARGETS FOR MULTI-DRUG RESISTANT TUBERCULOSIS THROUGH ANALYSIS OF NEXT GENERATION SEQUENCING DATA

M. Xavier Suresh*1,2 and S. Mira Priyadharshini2

1Department of Physics, Sathyabama University, Chennai -600119.
2Department of Bioinformatics, Sathyabama University, Chennai -600119.

Email: xaviersuresh@gmail.com

Received on 04-08-2016 Accepted on 25-08-2016

Abstract

Multi-Drug Resistant Tuberculosis (MDRTB) is an infectious form of tuberculosis that is resistant to two or more drugs. To combat this problem one specific approach is to find novel targets for finding new drugs. Therefore, at first, through literature and database search the basic drugs information, the specific targets, their associated genes are gathered and then the homologous proteins are found. Whether these proteins are present in the human proteome or not are checked and those not available are shortlisted. Docking studies have been carried out and binding sites were identified. These sites obtained are analysed with the NGS data for the find of novel targets, which revealed 15 novel targets.

Keywords: MDR-TB, tuberculosis, drug resistance, NGS, novel target, drug design.

Introduction

Tuberculosis is now, the major health problem in public. It is found to be the most infectious bacterial disease commonly affecting the lungs especially in the upper air sacs, caused by Mycobacterium tuberculosis (Mtb) which is a small and slow growing aerobic bacterium, that lives only in people [1]. Tuberculosis is found to be one of the three big infectious diseases killing approximately 1-2 million people worldwide noted from the survey taken by WHO Global Tuberculosis Report every year. Almost nine million people worldwide suffer from the active disease, among which, two billion i.e. around two-thirds of the population harbor the latent or the persistent infection of Mtb [2]. Moreover the major combination of the immunodeficiency virus along with Mycobacterium tuberculosis brings in the advent of anti-tubercular drug discoveries intense [3].

The manifestation of tuberculosis in general is of two categories, namely pulmonary and extra-pulmonary, in which, pulmonary is more often manifested and that, extra-pulmonary is occasionally manifested in terms of clinical...
tuberculosis. On a survey of immunocompetent patients, extra-pulmonary tuberculosis is found to constitute 15-20 percent of most of the clinical cases. The DOT therapy is an intensive drug treatment which was discovered by the WHO organisation for the relief of patients over continuous screening with the medication using few drugs from when the symptoms and surveillance of tuberculosis was discovered and reported. A continuous observation of this treatment has led to the success rates of a high survival in tuberculosis patients [4].

In recent years the development of drug-resistant strains of TB is a major threat. In order to combat MDR-TB there is a need for finding novel targets in invent of new drugs, however there is no proper high-throughput technology in the find of new targets. Raman et al 2008, proposed a target identification pipeline for finding novel targets for MDR-TB in which using flux balance analysis and network analysis, proteins critical for survival of *M. tuberculosis* are first identified, followed by comparative genomics with the host, finally incorporating a novel structural analysis of the binding sites to assess the feasibility of a protein as a target. The pipeline providing a useful platform leading to drug discovery, finding the targets on comparison of newly found targets with that of the ones available in the literature and validating them [5,6].

Isoniazid, rifampicin, ethambutol and pyrazinamide are used as the first-line drugs. The injectable drugs are, amikacin, capreomycin, veomycin and kanamycin. P-aminosalycilic acid, ethionamide and cycloserine, all being second-line bacteriostatics, though are very much established clinical efficacy, they have a most prominent side effects on the other hand on usage. The drugs namely isoniazid and ethionamide are the inhibitors of the mycolic acid synthesis [7-9]. The other set of drugs namely cycloserine and ethambutol are having the function as the inhibitors of the synthesis of peptidoglycan and cell wall arabinogalactan respectively thus weakening the bacteria’s cell wall [10]. Rifampin and amikacin show their pharmacological actions as the inhibitors of the bacterial RNA or protein synthesis [11,12].

There had been many ways of in silico processes of drug designing in terms of the drug and target identification. The target driven lead discovery leading to the identification of novel targets using some computational methods [13,14]. The steps to find the targets include the initial stage of the structural comparison of the different proteins to find the basic targets and the network interaction of the various proteins in a database called STRING, thus, forming a network of homologous proteins and justifying the homology with its score of 0.7 and more to identify the maximum homology as well as with a homology percentage criteria of 50% [15,6].
The recent advances in genome sequencing, the mutational analysis and the genomic landscapes brings into the venture of specifically the whole exome sequencing using NGS, thus gaining popularity in the existing technologies in terms of human genetic cases at moderate costs. It leads to a straight forward interpretation of manageable data amounts. This analysis includes a step of five set of distinct analytics including, the order of: assessment of quality, alignment, identification and annotation of variants and the last step being visualization using the different techniques of the NGS data analysis [16].

**Materials and Methods**

There are various materials used, which include the online tools and software’s such as, NCBI, UniProt, STRING, USEGALAXY, TB drug target database, DrugBank, PubChem, LigSite, Phyre2, PDB and other offline tools and software including, MEGA and DISCOVERYSTUDIO. The methodology, in general involves the retrieval of the drugs, their targets, finding of homologous proteins for each drug, checking the presence of these proteins in human proteome. Then, the shortlisted proteins are studied for interactions; the selected protein is docked against all the natural compounds including the plant compounds. The binding site information are extracted and are being used as the major input for the analysis of NGS data to screen and find the novel targets. The work flow is shown in fig 1.

![Workflow of NGS analysis](image)

**Fig 1: Workflow of NGS analysis.**

**Retrieval of NGS data**

The NGS data refers to the short reads of forward and reverse strands of the whole genome and whole exome sequencing. The NGS data that have been sequenced using the techniques as discussed earlier such as the Illumina,
Roche454, ion torrent and other forms of sequencing technologies is available online databases. The NCBI database that has the sequences of thousands of reads of the genomic and exomic fragments bringing in the speedest comparison of the protein sequences. The NGS data for TB is retrieved from the SRA section of the NCBI database. The sequences in each reads are compared with the sequences of the target proteins of each individual drug component already reported for the MDR-TB. This enable the comparison of the binding sites previously found in the MTB proteins with that of the NGS data to analyse and find novel target.

Selection of compounds and Differential binding analysis

The drug compounds and their chemical structures are retrieved from DrugBank. The drugs considered in this study are Rifampicin, Streptomycin, Rifabutin Ethambutol, Cycloserine, Pyrazinamide, Ofloxacin, P-aminosalicylic acid and Isoniazid. The pharmacophore properties, molecular descriptors were analysed and drugs that can bind with the target protein structures were studied using discovery studio. The predicted interactions of each drug with the target tuberculosis H37Rv proteins are utilised to map the interacting residues in the sequences using MEGA and this information is further used to screen the novel targets. Structures were modelled using Phyre2. Whether these proteins are present in the human proteome or not are checked and those not available are shortlisted.

A relatively new technique is the analysis of differential binding that draws much from the analysis of differential gene expression and has similar power to detect biologically meaningful binding changes between samples. The DiffBind package implemented in USEGALAXY allows the identification of genomic loci that are differentially bound between two conditions. It was developed around algorithms used for differential gene expression analysis by RNA-seq. These differential methods to predict the protein binding names that are functional enriched with reference annotation file. The resultant datasets are further used for protein structure prediction and validation.

Results and Discussion

The retrieval of the drugs along with their target and the basic function of each gene have been retrieved from the TB drug target database that provides a set of commonly known drugs for tuberculosis, which is especially used for MDR-TB analysis. After the find of the genes, their respective proteins were identified, their interacting proteins in the network were obtained from STRING db and checked and ensured that those are not present in the human proteome. Following this process, the proteins are shortlisted for 9 drugs. The selected drugs along with their targets are given in table1.
Table I: List of drugs and their respective targets

<table>
<thead>
<tr>
<th>SL. NO.</th>
<th>DRUG NAME</th>
<th>TARGET</th>
<th>GENE NAME</th>
<th>FUNCTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Isoniazid</td>
<td>KatG</td>
<td>Rv1908c</td>
<td>Catalase-peroxidase-peroxinitritase T katG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>InhA</td>
<td>Rv1484</td>
<td>Enoyl-(Acyl carrier protein) reductase</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DfrA</td>
<td>Rv2763c</td>
<td>Dihydrofolatereductasedf (dfr)</td>
</tr>
<tr>
<td>2</td>
<td>Rifampicin</td>
<td>RpoB</td>
<td>Rv0667</td>
<td>DNA-directed RNA polymerase beta sub unit</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RpoC</td>
<td>Rv0668</td>
<td>DNA-directed RNA polymerase beta' sub unit</td>
</tr>
<tr>
<td>3</td>
<td>Pyrazinamide</td>
<td>Fas</td>
<td>Rv2524c</td>
<td>Fatty acid synthetase</td>
</tr>
<tr>
<td>4</td>
<td>Streptomycin</td>
<td>RpsL</td>
<td>Rv0682</td>
<td>30S Ribosomal protein S12</td>
</tr>
<tr>
<td>5</td>
<td>Ethambutol</td>
<td>EmbC</td>
<td>Rv3793</td>
<td>Arabinosylindolylacetylinositol synthase</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EmbA</td>
<td>Rv3794</td>
<td>Arabinosylindolylacetylinositol synthase</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EmbB</td>
<td>Rv3795</td>
<td>Arabinosylindolylacetylinositol synthase</td>
</tr>
<tr>
<td>6</td>
<td>Cycloserine</td>
<td>Alr</td>
<td>Rv3423c</td>
<td>Analineracemase</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ddl</td>
<td>Rv2981c</td>
<td>D-alanylanlinesynthetase</td>
</tr>
<tr>
<td>7</td>
<td>P-aminosalicylic acid</td>
<td>FolK</td>
<td>Rv3606c</td>
<td>2-amino-4-hydroxy-6-hydroxymethylidihydropteridine pyrophosphokinase</td>
</tr>
<tr>
<td>8</td>
<td>Ofloxacin</td>
<td>GyrB</td>
<td>Rv0005</td>
<td>DNA topoisomerase IV-subunit B</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GyrA</td>
<td>Rv0006</td>
<td>DNA gyrase subunit A</td>
</tr>
<tr>
<td>9</td>
<td>Rifabutin</td>
<td>RpoA</td>
<td>Rv3457c</td>
<td>DNA-directed RNA polymerase alpha sub unit</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RpoB</td>
<td>Rv0667</td>
<td>DNA-directed RNA polymerase beta sub unit</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RpoC</td>
<td>Rv0668</td>
<td>DNA-directed RNA polymerase beta sub unit</td>
</tr>
</tbody>
</table>

The structures of these proteins were obtained from Protein Data Bank or modelled by homology modelling using Phyre 2. Further the protein structures are subjected to docking studies using the above said drugs in order to find out the ligand interaction sites at the best. For example, 4TRO protein target has been selected for isoniazid drug that shows very good interactions for the drug upon docking studies. All these sites have been considered for the find of novel targets using NGS data.

The NGS data for the most virulent strain H37Rv of Mtb was retrieved from the NCBI database. The list of ids are ERX651109, 651108, 8891, 8890, 8889, 8888, 8887, 8886, 8885, 8884, 8878, 8877, 8875, 8874, 8908, 8907, 8906, 8905, 8904, 8901, 8900, 8899, 8898, 8897, 8896, 8895, 8894. Analysis shows that there are novel targets compared with the binding site information for each drug. From the above data, novel targets have been found and additionally, for each protein has also been found.

Table II: Novel targets identified for each drug.

<table>
<thead>
<tr>
<th>DRUGS</th>
<th>NOVEL TARGETS</th>
</tr>
</thead>
</table>
Oflxacin
SLC22A5, SLC22A6, SLC22A2, SLC22A4, ABCC1, ABCC2, CYP3A4, CYP1A2, ABCB1, TOP2A

Cycloserine
MAB_0809c, MAB_0810c, MAB_0812, MAB_0811, MAB_4706c, MAB_4302

Pyrazinamide
pncA, katG, rpoB, gyrA, GyrN, rpsL, embB, embC, ahpC

PAS
NAT1, PLA2G2E, ALOX5, CHUK, PTGS2, MPO, NAT2

Isoniazid
NAT2, CYP2E1, CYP2C19, CYP3A4, CYP1A2, CYP2A6, CYP2B6, CYP2D6, CYP2C9, CYP2C8

Summary and Conclusion

Novel targets for MDRTB are identified through the analysis of the NGS data available in databases. The analysis process involves the combined techniques of sequence analysis and differential drug binding. The initial dataset with known targets is being used. Mapping the binding site onto the NGS data uncover several new genes, whose homologous proteins, then, performing the docking analysis results in novel targets. These targets could be further analysed and form the basis for finding novel drug molecules to fight MDR TB.

Acknowledgment: We thank Sathyabama University for the facilities provided to carry out the research work.

References


Corresponding Author:
M. Xavier Suresh*
Email: xaviersuresh@gmail.com