M. tuberculosis interactome analysis unravels potential pathways to drug resistance

Karthik Raman and Nagasuma Chandra*

Bioinformatics Centre and Supercomputer Education and Research Centre,

Indian Institute of Science, Bangalore – 560012

*** Corresponding author:**

Dr. Nagasuma Chandra

Bioinformatics Centre and Supercomputer Education and Research Centre

Indian Institute of Science

Bangalore - 560012

INDIA

Phone: +91-80-22932469

+91-80-22932892

Fax: +91-80-23600551

E-Mail: nchandra@serc.iisc.ernet.in

Abstract

Drug resistance is a major problem for combating tuberculosis. Lack of understanding of how resistance emerges in bacteria upon drug treatment limits our ability to counter resistance. By analysis of the *Mycobacterium tuberculosis* interactome network, along with drug-induced expression data from literature, we show possible pathways for the emergence of drug resistance. To a curated set of resistance related proteins, we have identified sets of high propensity paths from different drug targets. Many top paths were upregulated upon exposure to anti-tubercular drugs. Different targets appear to have different propensities for the four resistance mechanisms. Knowledge of important proteins in such pathways enables identification of appropriate *'co-targets'*, which when simultaneously inhibited with the intended target, is likely to help in combating drug resistance. RecA, Rv0823c, Rv0892 and DnaE1 were the best examples of co-targets for combating tuberculosis. This approach is also inherently generic, likely to significantly impact drug discovery.

Introduction

Tuberculosis (TB) has remained one of the largest killer infectious diseases despite the availability of several chemotherapeutic agents and a vaccine [1]. The global burden of tuberculosis has taken a new dimension in the recent years due to the emergence of drug resistant varieties of *Mycobacterium tuberculosis*, besides synergy with HIV [2]. Global surveillance indicates that multi-drug resistant (MDR-TB) and extensively drug resistant TB (XDR-TB) is spreading to many countries and is posing a major threat to TB eradication programs [3,4]. Several different strategies are being explored to counter the problem of resistance, which include rotation of antibiotic combinations, enhanced medical supervision to ensure patient compliance, identification of new targets that may be less mutable, search for new chemical entities for known targets, use of virulence factors as targets and 'phenotypic conversion', which aims to inhibit the resistance mechanism employed by the bacterium [5]. While each of these may be very important measures, available statistics indicate that resistant forms are still on the rise, warranting more research in the area. Of the different measures listed, the most cogent in its approach in the long term, is targeting the resistance mechanisms, since it enables confronting the problem at its source. However, in order to use this strategy effectively, it is at the outset, essential to understand the ways by which resistance can emerge upon exposure to a given drug.

Studies on the molecular mechanisms of resistance to first line and second line antitubercular drugs have led to mapping of several mutations in the drug targets and the regulatory gene segments [3]. Besides these, the activation of the efflux pumps and drug-modifying enzymes are other known mechanisms of drug resistance [6]. Several studies in other organisms have reported the acquisition of drug-inactivating genes through horizontal gene transfer as a means of selection of the resistant variety [7]. It is clear from these that diverse mechanisms can exist for generating resistance and that the proteins involved in each can be quite remote from the drug targets in terms of their functional classes. It is important to study how, upon exposure to the specific drug molecule, information can flow from the target molecule to those that are involved in resistance. Knowledge of the molecular basis by which information flows from the specific drug target to the proteins elsewhere in the system relevant to drug resistance,

will help us address the issue of resistance in more systematic, rational and novel ways. With the availability of many genome-scale data from several studies, it is now feasible to address the issue of resistance from a systems perspective. Here, we use a proteome-scale network of protein–protein connectivity to discover possible pathways that may be responsible for generating drug resistance. The network analyses reported here further help in classification of these paths based on known resistance mechanisms. The study also identifies controlling hubs within these paths and suggests proteins that could be explored for their use as drug co-targets.

Results and Discussion

Interactome network. A proteome-scale interaction network of proteins in *M. tuberculosis H37Rv (Mtb)* was derived (Table S1) *from* the STRING database [8], which includes interactions from published literature describing experimentally studied interactions as well as those from genome analysis using several well established methods such as domain fusion, phylogenetic profiling and gene neighborhood concepts. Thus, the network captures different types of interactions such as (a) physical complex formation between two proteins required to form a functional unit, (b) genes belonging to a single operon or to a common neighborhood, (c) proteins in a given metabolic pathway and hence influenced by each other, (d) proteins whose associations are suggested based on predominant co-existence, co-expressions, or domain fusion. This network represents a first comprehensive view of the connectivity among the various proteins, analogous to obtaining the road map of a city. With the methodology currently available, it is inevitable for a network of this type to contain some false positives as well as false negatives. To minimize this problem, all interactions tagged as 'low-confidence' in the STRING database have been eliminated from this study. The considered network, despite its shortcomings, provides an excellent framework for navigating through the proteome. It also allows for refinement of the network upon the availability of new experimental data.

Paths to Resistance. The network contains 3,925 nodes (proteins), with 29,664 undirected edges (interactions) between them. The clustering coefficient of the network is 0.447. This high clustering coefficient is indicative of the high density of connections

in the network. Global views of the effects of antitubercular drugs on the mycobacterial proteome were obtained from the microarray experiments reported earlier; a list of genes whose expression levels were either increased or decreased upon exposure to the drugs was obtained for seven drugs, *viz.* isoniazid [9,10], ethionamide [10], isoxyl, tetrahydrolipstatin, SRI-221, SRI-967 and SRI-9190 [9]. Known mechanisms relevant to resistance were classified into four types (a) efflux pumps, which transport drugs out of the cell, (b) cytochromes and other target modifying enzymes that could cause potential chemical modification of drug molecules, (c) SOS-response and DNA replication leading to mutations in the gene or its regulatory region, (d) proteins involved in horizontal gene transfer (HGT) to import a target modifying or detoxifying protein from its environment. We curated a list of 74 genes in the *Mtb* genome based on these mechanisms from published literature (Table 1). Among these, 12 cytochromes, out of the 36 present, five pumps (of 25), one HGT protein (of four) and three SOS proteins (of nine) were found to be upregulated in the expression profile corresponding to at least one of the seven drugs considered here.

Proteins that were targets of the given drug(s) were grouped together as '*source*' while the curated set of resistance proteins were grouped together as '*sink*'. Three of the drugs isoniazid, ethionamide and isoxyl are known to be inhibitors of mycolic acid biosynthesis, for which the 26 proteins of the mycolic acid pathway (MAP) [11] were used as *source*. It can be envisaged that upon inhibition of a protein in a given pathway, metabolic adjustments often occur so as to minimize the effect of inhibition on the particular protein [6]. In order to incorporate the effect of such adjustments, we have considered the whole pathway as the *source* rather than the individual protein. There are also reports in the literature that multiple proteins in MAP may be targeted by some of these drugs [12], making it important to consider the pathway as a whole. Shortest paths from *source* to the curated set of resistance proteins (*sink*) were computed. The set of shortest paths forms a network, consisting of 616 nodes and 1683 edges, which is referred to as the MAP-RES network hereafter (Fig. 1).

The best ranked shortest paths to each of the four resistance mechanisms were identified as shown in Table 2 (complete list given in Table S2). High-scoring paths to each of the resistance mechanisms were observed from MAP (Fig. 2). However, paths

to SOS proteins top the list, followed by paths to cytochromes, while paths to HGT and pumps were of much lower rank. Nodes and edges that occur most frequently in a given set of paths are considered as node and edge hubs. Top edge-hubs in MAP-RES are listed in Table S3. Several of the nodes in these edge hubs also happen to be top node hubs. Some such edge hubs are MmaA4 (Rv0642c) – Rv0892, FabD (Rv2243) – FadA5 (Rv3546), FadA5 (Rv3546) – Cyp125 (Rv3545c), Rv0049 – PcaA (Rv0470c), Rv0823c – DesA1 (Rv0824c), Acs (Rv3667) – Rv3779 and KasA (Rv2245) – RecA (Rv2737c).

It is interesting to note that many of these edges can be attributed to metabolic linkages wherein the reactions involving the two proteins share a common metabolite (e.g. Rv0642 – Rv0892 or Rv0892 – Rv3801c). A few other edges could be attributed to adjacency of the genes in the genome whose transcription may be regulated by a common mechanism. A number of these proteins are also upregulated in the top paths (Figs. 1 & 2). In some cases, the entire paths were upregulated in one or more of the drugs, indicating the correlation of the identified paths with the observed expression profiles. A path from FabG4 (Rv0242c) to KasA (Rv2245) to RecA (Rv2737c), appears to be such a path, where all are upregulated, making the flow of information from *source* to *sink* that of high propensity. Another interesting path is from DesA1 (Rv0824c)–Rv0823c–RecA (Rv2737c), which also has an alternate sub-path to RecA through RuvA (Rv2593c). In this path, besides RecA, Rv0823c is also upregulated (Fig. 2). Considering the individual functions of these molecules, it is easy to comprehend that the transcriptional regulator (Rv0823c) influenced by DesA1 (Rv0824c), triggers the activation of RecA (Rv2737c), which while itself a *sink*, also activates many proteins such as DnaE1 (Rv1547), important for DNA synthesis. DesA1 was earlier identified as a potential anti-tubercular drug target [11], by virtue of its critical role in mycolic acid biosynthesis [13]. Targeting RecA, DnaE1 or Rv0823c individually or together by chemotherapeutic agents along with DesA1, appears to be a good strategy to counter emergence of drug resistance. Thus RecA and DnaE1 could be considered as potential ‘co-targets’ for DesA1 and other MAP targets. All these three proteins are upregulated in response to one or more of the MAP inhibitors.

As many as 19 cytochrome proteins were found to be present in a tight cluster (Fig. 1), seven of them upregulated, connected to the MAP through Rv0892, which is annotated as a probable monooxygenase. Some noteworthy pathways across the four classes are: (i) Rv2243 (FabD) – Rv3546 (FadA5) – Rv3545c (Cyp125) (cytochrome); (ii) Rv2524c (Fas) – Rv2918c (GlnD) – Rv2890c (RpsB) – Rv3240c, the latter a translocase implicated in horizontal gene transfer; (iii) Rv2245 (KasA) – Rv1908c (KatG) – Rv1988 (a methyltransferase annotated as an efflux pump); (iv) Rv2245 (KasA) – Rv2737c (RecA), Rv0904c (AccD3) – Rv1547 (DnaE1), both important for homologous recombination and DNA synthesis.

In fact, RecA (Rv2737c) and a few other proteins such as SecA1 (Rv3240c), SahH (Rv3248c), Rv0892 and metK (Rv1392) occur in multiple resistance mechanisms, making their roles even more prominent in the emergence of resistance. Some of the MAP proteins such as KasA (Rv2245), FabD (Rv2243), which are upregulated, also appear to mediate multiple pathways across different resistance mechanisms. Proteins important for multiple resistance pathways could help in prioritizing 'co-targets'. Of the top node and edge hubs in MAP-RES, it is of interest to note that nodes Rv0892 and Rv2243 are also among the top hubs in the entire STRING network, indicating their critical role in the *Mtb* interactome. Table S4 lists the nodes and edges in the STRING, ranked based on their betweenness.

Table 2 also indicates shortest paths containing upregulated proteins in one or more of the four non-MAP drugs, used for obtaining the microarray data, along with the three MAP drugs discussed so far. It was observed that while paths to SOS still topped the list, many paths to cytochromes, pumps and HGT were found and contained a number of upregulated proteins, indicating that propensities for traversal of a path could vary from drug to drug. The other known targets of clinically used anti-tubercular drugs are RpoB (Rv0667) for rifampicin, the gyrase GyrA (Rv0006) for fluoroquinolones, cell wall biosynthesis proteins such as Alr (Rv3423c), DdlA (Rv2981c), Rv3792, EmbA (Rv3794) and EmbB (Rv3795) for cycloserine. With these targets as *source*, shortest paths were computed to the *sink* proteins. We observe that short high-scoring paths were observed from all targets examined to SOS involving many common nodes (e.g. DnaE1

(Rv1547), RecA (Rv2737)) whereas paths to other resistance mechanisms differed from target to target. Some edges in the SOS response (e.g. Rv2158c – Rv0631c) were common to paths from cell wall proteins and gyrase. From gyrase, SOS was the only predominant path, whereas from RpoB, paths to HGT and cytochromes were also among the top paths. For RpoB, many of the paths to SOS were mediated through Dcd (Rv0321), a DCTP deaminase. From cell wall proteins, paths to HGT and SOS top the list of high-scoring paths again, although there are also several paths to cytochromes. Paths to pumps were relatively fewer in number, in all cases. Of interest, however was a path to EfpA (Rv2846c), a transporter known to confer resistance to fluoroquinolones (Rv0006–Rv0524–Rv3065–Rv2846c), rifampicin and isoniazid resistance [6]. A higher scoring path to IniA (Rv0342) was observed through Rv0340 (conserved hypothetical protein), again agreeing well with previous reports based on transcription studies [14]. We also analyzed paths from proteins KatG (Rv1908c), EthA (Rv3854c) and PncA (Rv2043c), all known to transform prodrugs to the active drug species to the *sink*. It was interesting to observe that while EthA and PncA did not appear to have short paths to any of the resistance mechanisms, KatG had a direct interaction with an efflux pump (Rv1988, a probable methyltransferase, homologue of Erm37) and also several paths of two or fewer edges to all resistance mechanisms. In fact, studies on clinical isolates have shown mechanisms such as mutations in KatG, leading to loss of catalase activity, mutations in the promoter region of *inhA*, leading to its overexpression and mutations in InhA, leading to loss of affinity for isoniazid [3]. Paths to different resistance mechanisms for different drugs observed here, suggest that a given target may have a higher propensity for eliciting a specific mechanism of resistance, which when understood can be utilized to identify appropriate co-targets or secondary targets, to prevent emergence of resistance.

Upon analysis of the network formed only by the upregulated proteins (Table S5), for each of the seven drugs, we observe that the network density is much higher, over an order of magnitude in most cases, than that of the whole interactome. This indicates that the upregulated genes have a higher influence on each other and more importantly their inter-relatedness is perhaps for a specific purpose. Such a purpose, if present, is non-obvious by their functional classes or by a common analysis (such as co-expression) of

the microarray data, except in a few cases. However, when viewed in the context of the interaction networks, it becomes possible to analyze the outcome of such inter-relatedness.

Reports in literature indicate that the main strategies for adaptation by the bacillus in response to these drugs are mutations to reduce the bioavailability or the binding of these drugs to the MAP target(s). This inherently implies that the MAP proteins have a streamlined mechanism to pass on the information of drug inhibition (or lack of mycolic acid production) to the SOS proteins involved in recombination and DNA biosynthesis. A time course microarray data over a period of a few weeks from clinical samples would ideally have been required to see the expression patterns where resistance has emerged. Nevertheless, the existing data provides a first glimpse of the possible routes that might lead to resistance. Knowledge of the pathways leading to drug resistance will be of immense help in identifying appropriate co-targets, thus forming a new rational strategy for combating drug resistance. Given the rapid accumulation of various types of '*omics*' data, including comprehensive views of protein–protein interactions, this type of analyses is becoming feasible for many pathogenic organisms. Our approach is also inherently generic, lending itself to be utilized in any drug discovery program.

Methods

Interactome Network. A proteome-scale interaction network of proteins in *Mtb* was derived (Table S1) from the STRING database, using only the 'high-confidence' and 'medium-confidence' data.

Curation of Resistance Genes. We curated a list of 74 genes in the *Mtb* genome based on resistance mechanisms from published literature [6,7,15–17, TubercuList web server (http://genolist.pasteur.fr/TubercuList/)] (Table 1).

Betweenness. Betweenness is a centrality measure of a vertex within a graph [18]. For a graph **G(V,E)**, with *n* vertices, the betweenness $C_B(v)$ of a vertex *v* is defined as

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

where σ_{st} is the number of shortest paths from *s* to *t*, and $\sigma_{st}(v)$ is the number of shortest paths from *s* to *t* that pass through a vertex *v*. A similar definition for 'edge betweenness' was given by Girvan and Newman [19]. Betweenness was calculated for all nodes and edges in STRING, ranking them to obtain a list of hubs (Table S4).

Network analysis. Shortest paths from *source* to the curated set of resistance proteins (*sink*) were computed using the Dijkstra's algorithm implemented in the MATLAB-Boost Graph Library (David Gleich; http://www.stanford.edu/~dgleich/programs/matlab_bgl/). We added a weighting scheme to account for the frequency of the edge in the network, as well as to incorporate the upregulation information of the nodes forming a given edge. The weight of an edge *AB* is given as

$$w_{st} = \frac{1}{f_{st}(1+N_s)(1+N_t)}$$

where f_{st} corresponds to the frequency of the edge between *s* and *t*, which is the number of times a given edge occurs in the set of paths, N_s and N_t refer to number of drugs (only drugs that target MAP are considered) for which node *s* and node *t* are upregulated respectively. Thus, the maximum possible score for an edge is unity, when the edge appears only once in the set of chosen paths and neither of the nodes forming the edge is upregulated. A path score was computed as the sum of the weighted scores for the edges in a path, from which the least scoring path corresponded to the highest rank. Cytoscape [20] was used for network visualizations.

Acknowledgements. Financial support from the Mathematical Biology initiative of the Department of Science and Technology, Govt. of India, is gratefully acknowledged. Use of facilities at the Interactive Graphics Based Molecular Modeling Facility and Distributed Information Centre (both supported by the Department of Biotechnology) and the facilities at the Supercomputer Education and Research Centre are also gratefully acknowledged.

Competing Interests. The authors declare that no competing interests exist.

Author Contributions. NC generated the idea. KR and NC designed the experiments. KR carried out the computations. KR and NC analyzed the data and wrote the manuscript.

References

1. World Health Organisation. 2007 tuberculosis facts. Tech. rep., WHO (2007).
2. Nunn, P., Williams, B., Floyd, K., Dye, C., Elzinga, G. *et al*. Tuberculosis control in the era of HIV. *Nat. Rev. Immunol.* **5**:819–826 (2005).
3. Johnson, R., Streicher, E. M., Louw, G. E., Warren, R. M., van Helden, P. D. *et al*. Drug Resistance in *Mycobacterium tuberculosis*. *Curr. Issues Mol. Biol.* **8**:97–112 (2006).
4. World Health Organisation. Emergence of XDR-TB. WHO concern over extensive drug resistant TB strains that are virtually untreatable. Available: http://www.who.int/mediacentre/news/notes/2006/np23/en/index.html (2006).
5. Tan, Y.-T., Tillett, D. J. & McKay, I.A. Molecular strategies for overcoming antibiotic resistance in bacteria. *Mol. Med. Today* **6**:309–314 (2000).
6. Nguyen, L. & Thompson, C. J. Foundations of antibiotic resistance in bacterial physiology: the mycobacterial paradigm. *Trends Microbiol.* **14**:304–312 (2006).
7. Smith, P. A. & Romesberg, F. E. Combating bacteria and drug resistance by inhibiting mechanisms of persistence and adaptation. *Nat. Chem. Biol.* **3**:549–556 (2007).
8. von Mering, C., Jensen, L. J., Kuhn, M., Chaffron, S., Doerks, T. et al. STRING 7 – recent developments in the integration and prediction of protein interactions. *Nucleic Acids Res.* **35**:358–362 (2007)
9. Waddell, S. J., Stabler, R. A., Laing, K., Kremer L, Reynolds R. C. *et al*. The use of microarray analysis to determine the gene expression profiles of *Mycobacterium*

tuberculosis in response to anti-bacterial compounds. *Tuberculosis* **84**:263–274 (2004).

10. Fu, L. M. Exploring drug action on *Mycobacterium tuberculosis* using affymetrix oligonucleotide genechips. *Tuberculosis (Edinb)* **86**:134–143 (2006).

11. Raman, K., Rajagopalan, P. & Chandra, N. Flux balance analysis of mycolic acid pathway: Targets for antitubercular drugs. *PLoS Comp. Biol.* **1**:e46 (2005).

12. Mdluli, K., Slayden, R. A., Zhu Y., Ramaswamy S., Pan, X. *et al*. Inhibition of a *Mycobacterium tuberculosis* beta-ketoacyl acyl carrier protein synthase by isoniazid. *Science* **280**:1607–1610 (1998).

13. Raman, K., Rajagopalan, P. & Chandra, N. Hallmarks of mycolic acid biosynthesis: A comparative genomic study. *Proteins.* **69**:358–368 (2007).

14. Colangeli, R., Helb. D., Sridharan, S., Sun, J., Varma-Basil, M. *et al*. The *Mycobacterium tuberculosis iniA* gene is essential for activity of an efflux pump that confers drug tolerance to both isoniazid and ethambutol. *Mol. Microbiol.* **55**:1829–1840 (2005).

15. Rossi, E. D., Arrigo, P., Bellinzoni, M., Silva, P. A. E., Martín, C. *et al*. The multidrug transporters belonging to major facilitator superfamily in *Mycobacterium tuberculosis*. *Mol. Med.* **8**:714–724 (2002).

16. Buriánková, K., Doucet-Populaire, F., Dorson, O., Gondran, A., Ghnassia, J. C. *et al*. Molecular basis of intrinsic macrolide resistance in the *Mycobacterium tuberculosis* complex. *Antimicrob. Agents. Chemother.* **48**:143–150 (2004).

17. Madsen, C. T., Jakobsen, L., Buriánková, K., Doucet-Populaire, F., Pernodet, J. L. *et al*. Methyltransferase Erm(37) slips on rRNA to confer atypical resistance in *Mycobacterium tuberculosis*. *J Biol. Chem.* **280**:38942–38947 (2005).

18. Freeman, L. C. A set of measures of centrality based on betweenness. *Sociometry* **40**:35–41 (1977).

19. Girvan, M. & Newman, M. E. J. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. U S A* **99**:7821–7826 (2002).

20. Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T. *et al*. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome. Res.* **13**:2498–2504 (2003).

Antibiotic efflux pumps [6,15]

PstB (Rv0933), Rv2686c, Rv2687c, Rv1688c, IniA (Rv0342), Mmr (Rv3065), Rv3239c, Rv3728, EfpA (Rv2846c), Rv1877, Rv2333c, Rv2459, Rv1410c, Rv1250, Rv1258c, EmrB (Rv0783c), Rv1634, Rv0849
Rv0191, Rv0037c, Rv2456c, Rv2994 (hypothetical) [15]

BlaC (Rv2068c) [Antibiotic degrading enzyme [**6**]]*

Erm37 (Rv1988), WhiB7 (Rv3197A) [Target-modifying enzymes [16,17]]*

SOS and related genes [7]

DnaE2 (Rv3370c), RuvA (Rv2593c), RecA (Rv2737c), RecB (Rv0630c), RecC (Rv0631c), RecD (Rv0629c), DnaE1 (Rv1547), PolA (Rv1629), LexA (Rv2720)

Genes implicated in horizontal gene transfer [TubercuList,7]

SecA1 (Rv3240c), SecA2 (Rv1821), Rv3659c, Rv3660c

Cytochromes [TubercuList]

ccdA (Rv0527), CcsA (Rv0529), CtaB (Rv1451), CtaC (Rv2200c), CtaD (Rv3043c), CtaE (Rv2193), CydA (Rv1623c), CydB (Rv1622c), CydC (Rv1620c), CydD (Rv1621c), Cyp121 (Rv2276), Cyp123 (Rv0766c), Cyp124 (Rv2266), Cyp125 (Rv3545c), Cyp126 (Rv0778), Cyp128 (Rv2268c), Cyp130 (Rv1256c), Cyp132 (Rv1394c), Cyp135A1 (Rv0327c), Cyp135B1 (Rv0568), Cyp136 (Rv3059), Cyp137 (Rv3685c), Cyp138 (Rv0136), Cyp139 (Rv1666c), Cyp140 (Rv1880c), Cyp141 (Rv3121), Cyp142 (Rv3518c), Cyp143 (Rv1785c), Cyp144 (Rv1777), Cyp51 (Rv0764c), DipZ (Rv2874), LldD1 (Rv0694), LldD2 (Rv1872c), QcrB (Rv2196), QcrC (Rv2194), SdhC (Rv3316)

Table 1: Curated list of resistance proteins

*These proteins are antibiotic degrading and target modifying proteins, but since they are very few, have been included with the pumps for convenience

Path	Score
SOS	
Rv0242c -(276)-> **Rv2245** -(596)-> **Rv2737c**	0.0102
Rv0242c -(276)-> **Rv2245** -(596)-> **Rv2737c -(44)-> Rv2720**	0.0727
Rv0904c -(241)-> **Rv1547**	0.1000
Rv0824c -(162)-> **Rv0823c** -(247)-> **Rv2737c**	0.1344
Rv0242c -(276)-> **Rv2245** -(596)-> **Rv2737c -(96)-> Rv2593c**	0.1352
Rv2246 -(400)-> **Rv1131** -(526)-> Rv1629	0.1389
CYTOCHROMES	
Rv2243 -(200)-> Rv3546 -(395)-> Rv3545c	0.0174
Rv0242c -(436)-> Rv2243 -(200)-> Rv3546 -(395)-> Rv3545c	0.0203
Rv1350 -(205)-> Rv2243 -(200)-> Rv3546 -(395)-> Rv3545c	0.0230
Rv1483 -(121)-> Rv2243 -(200)-> Rv3546 -(395)-> Rv3545c	0.0352
Rv2243 -(257)-> Rv0769 -(558)-> Rv0766c	0.0643
Rv0242c -(436)-> **Rv2243** -(559)-> Rv2782c -(513)-> **Rv1622c**	0.0655
Rv0242c -(436)-> Rv2243 -(257)-> Rv0769 -(558)-> Rv0766c	0.0673
Rv0242c -(436)-> **Rv2243** -(559)-> Rv2782c -(578)-> **Rv2193**	0.0738
Rv0643c -(350)-> Rv0892 -(350)-> **Rv0568**	0.0956
Rv0643c -(350)-> Rv0892 -(350)-> **Rv3059**	0.0956
ANTIBIOTIC EFFLUX PUMPS	
Rv2245 -(565)-> Rv0340 -(447)-> **Rv0342**	0.0526
Rv0242c -(276)-> Rv2245 -(565)-> Rv0340 -(447)-> Rv0342	0.0580
Rv2243 -(356)-> Rv2238c -(569)-> Rv2687c	0.0901
Rv2243 -(2)-> **Rv2245** -(565)-> Rv0340 -(447)-> **Rv0342**	0.1151
Rv0642c -(350)-> **Rv3248c** -(486)-> **Rv3240c** -(323)-> Rv3239c	0.1365
Rv0642c -(350)-> **Rv3248c** -(350)-> **Rv1988**	0.1385

Rv1350 -(90)-> **Rv2245** -(596)-> **Rv2737c** -(429)-> **Rv2882c** -(584)-> Rv0783c	0.3169
Rv2245 -(369)-> **Rv1908c** -(350)-> **Rv1988**	0.3833
Rv0242c -(276)-> Rv2245 -(369)-> **Rv1908c** -(350)-> **Rv1988**	0.3888
HORIZONTAL GENE TRANSFER	
Rv0644c -(350)-> **Rv3248c** -(486)-> **Rv3240c**	0.0812
Rv2243 -(202)-> Rv2925c -(378)-> Rv3659c	0.0913
Rv2243 -(202)-> Rv2925c -(378)-> Rv3659c -(63)-> Rv3660c	0.1538
Rv2245 -(596)-> **Rv2737c** -(514)-> **Rv2890c** -(187)-> **Rv3240c**	0.3715
Rv2524c -(464)-> **Rv2918c** -(380)-> **Rv2890c** -(187)-> **Rv3240c**	0.4833

Table 2: Top paths in the MAP-RES network.

The weighted score for each path is shown. The nodes upregulated by MAP drugs are indicated in black boldface, while those that are upregulated by other drugs are indicated in grey boldface. The figures in parentheses indicate the edge weight; a lower edge weight indicates a higher confidence in the interaction.

Figure legends:

Fig. 1: Illustration of the MAP-RES network depicting paths from MAP proteins to the curated set of resistance proteins. Nodes correspond to the individual proteins in the network while the edges indicate interactions between them. Each class of nodes is colored differently as indicated. Grey nodes indicate those that do not belong to any of the marked classes. The nodes are sized in proportion to the number of MAP drugs that induce its upregulation. The thickness of the edge is proportional to the number of times a shortest path is traversed through that edge.

Fig. 2: Top scoring paths from MAP to each of the four resistance classes are shown. Nodes are labeled by their Rv IDs, as obtained from TubercuList. Nodes correspond to the individual proteins in the network while the edges indicate interactions between them. Each class of nodes is colored differently as indicated. Grey nodes indicate those that do not belong to any of the marked classes. The nodes are sized in proportion to the number of MAP drugs that induce its upregulation. The thickness of the edge is proportional to the number of times a shortest path is traversed through that edge. The dotted edge is not a high-scoring path but is of significance, as discussed in the text.