

## RESEARCH ARTICLE

## Open Access



# Prediction of host - pathogen protein interactions between *Mycobacterium tuberculosis* and *Homo sapiens* using sequence motifs

Tong Huo<sup>1,2,4</sup>, Wei Liu<sup>1,2,4</sup>, Yu Guo<sup>1,3,4</sup>, Cheng Yang<sup>1,3,4</sup>, Jianping Lin<sup>1,3,4\*</sup> and Zihao Rao<sup>1,2,4</sup>

## Abstract

**Background:** Emergence of multiple drug resistant strains of *M. tuberculosis* (MDR-TB) threatens to derail global efforts aimed at reigning in the pathogen. Co-infections of *M. tuberculosis* with HIV are difficult to treat. To counter these new challenges, it is essential to study the interactions between *M. tuberculosis* and the host to learn how these bacteria cause disease.

**Results:** We report a systematic flow to predict the host pathogen interactions (HPIs) between *M. tuberculosis* and *Homo sapiens* based on sequence motifs. First, protein sequences were used as initial input for identifying the HPIs by 'interolog' method. HPIs were further filtered by prediction of domain-domain interactions (DDIs). Functional annotations of protein and publicly available experimental results were applied to filter the remaining HPIs. Using such a strategy, 118 pairs of HPIs were identified, which involve 43 proteins from *M. tuberculosis* and 48 proteins from *Homo sapiens*. A biological interaction network between *M. tuberculosis* and *Homo sapiens* was then constructed using the predicted inter- and intra-species interactions based on the 118 pairs of HPIs. Finally, a web accessible database named PATH (Protein interactions of *M. tuberculosis* and Human) was constructed to store these predicted interactions and proteins.

**Conclusions:** This interaction network will facilitate the research on host-pathogen protein-protein interactions, and may throw light on how *M. tuberculosis* interacts with its host.

## Background

Tuberculosis (TB), caused by *Mycobacterium tuberculosis* (MTB), is a major global health concern [1]. According to the World Health Organization (WHO) report [2], there were an estimated 8.7 million new cases of TB (13% co-infected with HIV) and 1.4 million TB-related deaths in 2011. Clearly, the number of TB-related deaths in single year is alarmingly higher than the roughly 300,000 deaths reported for the bird flu pandemic in 2009 [3]. Further, the regimens recommended for the treatment of TB are complex, often very long and include highly toxic drugs that have side effects. An antibiotic course consisting of

four first-line drugs like isoniazid, rifampicin, ethambutol and pyrazinamide for six months is recommended for treatment of TB. These first-line drugs were discovered more than 50 years ago [2,4]. Drug discovery for TB continues to lag behind. Co-infection with retroviruses like HIV further complicates TB treatment. Emergence of multi-drug resistant and extensively-drug resistant strains of *Mycobacterium* has threatened to derail global efforts for reigning in this pathogen [5]. Therefore, there is an urgent need to develop new anti-mycobacterial drugs [4] through an understanding of the genetics and physiology of *M. tuberculosis*.

*M. tuberculosis* primarily infects the respiratory system where it encounters alveolar macrophages and dendritic cells patrolling the lungs. However, the bacterium has an uncanny ability to survive the onslaught and in fact it uses the host macrophages for replication [5]. Virulence

\* Correspondence: [jianpinglin@nankai.edu.cn](mailto:jianpinglin@nankai.edu.cn)

<sup>1</sup>State Key Laboratory of Medicinal Chemical Biology, Nankai University, Tianjin 300071, China

<sup>3</sup>College of Pharmacy, Nankai University, Tianjin 300071, China

Full list of author information is available at the end of the article

factors like an unusual cell wall made up of mycolic acid, *UreC* gene that prevents acidification of phagosomes, and the ability of the pathogen to neutralize reactive nitrogen and oxygen intermediates using reductases helps the bacterium evade the host immune system. In addition to macrophages, T-cells have been shown to participate in host cell response against mycobacterium [6,7]. However, mycobacterium evades elimination by the host immune response and causes disease. Therefore, it is essential to study the interactions between *M. tuberculosis* and the host to learn how these bacteria cause disease [8]. The availability of the complete genome sequence of the pathogen *M. tuberculosis* [1] and the host *Homo sapiens* [9] provides an essential tool for prediction of these host-pathogen protein interactions.

Host-pathogen protein interactions (HPIs) are often involved in the pathogen's strategy to invade the host organism, breach the host's immune defenses, as well as replicate and persist within the organism [10,11]. Experimentally, there are two main approaches for detecting interacting proteins: binary approaches such as the yeast two-hybrid (Y2H) system and luminescence-based mammalian interactome mapping and co-complex methods such as co-immunoprecipitation (coIP) coupled with mass spectrometry (MS) [12]. However, these methods are time-consuming and expensive, especially when adopted in high-throughput mode [13]. Therefore, many computational methods have been developed to improve the coverage, accuracy, and efficiency in identifying protein pairs. These methods for predicting protein-protein interaction (PPI) take advantages of high-throughput data [14] and are based on protein sequence, structural and genomic features that are related to interactions and functional relationships [15,16], including phylogenetic profile [17,18], gene neighbor and gene cluster methods [19,20] and interologs [21,22]. Interologs, also referred to as homologous PPI method, is based on the assumption that homologous proteins preserve their ability to interact [23]. Recently, it has been applied for not only recognizing PPIs within an individual organism [24,25], but has also been used to detect host-pathogen protein interactions [26,27].

In this work, we developed a systematic flow to predict the HPIs between *M. tuberculosis* and *Homo sapiens* based on sequence motifs. First, protein sequences were used as initial input for identifying the HPIs between *M. tuberculosis* and *Homo sapiens* by 'interolog' method. The HPIs were further filtered by domain-domain interactions (DDIs) prediction. Then, protein functional annotations and existing experiments results were applied to remaining HPIs. As a result, 118 pairs of HPI were identified, which involve 43 proteins from *M. tuberculosis* and 48 proteins from *Homo sapiens*. Intra-species PPIs were further predicted for the proteins from *M. tuberculosis*

and proteins from *Homo sapiens* using VisANT [28], Reactome [29], InteroPorc [30], IntAct [31], DIP [32], MPIDB [33], MINT [34], and HPRD [35]. A biological interaction network between *M. tuberculosis* and *Homo sapiens* was then constructed by the predicted inter- and intra-species interactions. Finally, a database named PATH (Protein interactions of *M.tuberculosis* and Human) was constructed to store these predicted interactions and proteins.

## Methods

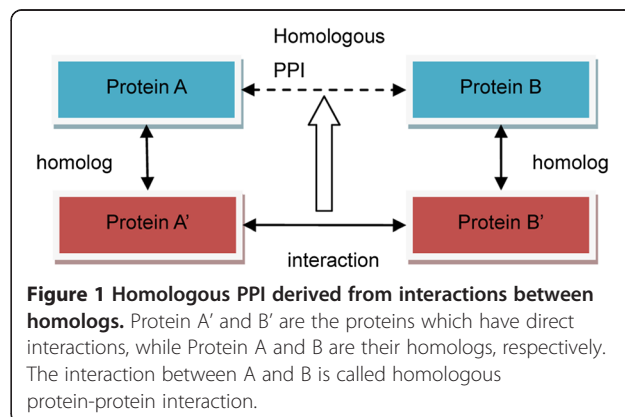
### Identifying HPIs by sequence comparison

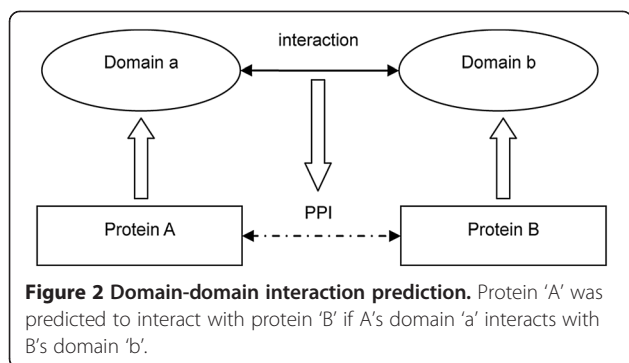
Figure 1 shows the procedure used to identify HPIs. The procedure was based on the rationale underlying interolog [36], which implies that two proteins (A and B) are predicted to interact if their relative homologs (A' and B') interact.

To predict homologs, the basic local alignment search tool BLAST (Basic local alignment search tool) [37] was employed to compute sequence similarities. Query protein sequences were aligned against all sequences with known interactions stored in the databases BIPS [38] and HPIDB [39]. BIPS and HPIDB are integrated databases including several data sources such as DIP [32] and IntAct [31], and both of the databases allow the users to set the parameters freely. The e-value and identity parameters were set to 1e-10 and 30 respectively, and the source of target interactors was set to *Homo sapiens* (taxid:9606). The query protein sequences were obtained from TB database [40].

### Detecting domain-domain interactions (DDIs)

Domains play an important role in mediating protein-protein interactions [41,42]. The studies on DDI (domain-domain interaction) are based on the assumptions that: (1) DDIs are independent of each other, and (2) two proteins interact if at least one pair of domains from two proteins interacts. DDI were constructed in three steps - 1) the protein sequences of *Homo sapiens* and *Mycobacterium tuberculosis* were assigned to families or domains; 2) a whole domain-domain interaction network was drawn; 3) mapping the domain 'a' from protein 'A' in *Homo sapiens* and





domain 'b' from protein 'B' in *Mycobacterium tuberculosis* to the whole network. If the domain 'a' interacted with domain 'b', the protein 'A' was predicted to interact with the protein 'B' (as in Figure 2).

To identify DDIs, the proteome of *M. tuberculosis* and *Homo sapiens* was aligned with Pfam families or domains with an E-value cut-off of  $1e-10$  using the Pfam-map program [43]. Then, protein-domain databases including 3DID [44], iPfam [45], DOMINE [46], DAPID [47] were selected to draw the DDI map.

#### Filtering HPIs by biological context or functional annotation

The information of each protein in the HPI pairs (including subcellular location, tissue specificity, biological process, molecular function, and cellular component) was obtained from the Uniprot website (www.uniprot.org). If the functional annotation of the pair of interactors in the quasi-credible HPI was found to correspond with at least one of the defined terms, the quasi-credible HPI was selected and upgraded as credible HPIs. The terms were selected from previously published studies on the infection and pathology of MTB [48-51].

#### Identifying intraspecific PPI network in *Homo sapiens* and *Mycobacterium tuberculosis*

The protein A from *Homo sapiens*, and protein B from *Mycobacterium tuberculosis*, that were involved in a HPI, were further screened against PPI databases to identify intraspecific PPIs. The resource of the intraspecific PPI for *Mycobacterium tuberculosis* included VisANT, Reactome, InteroPorc, IntAct, DIP, MPIDB, MINT, whereas for *Homo sapiens*, IntAct, HPRD, MINT, Reactome, DIP were included. We also attempted to use more databases such as virusmint [52], virhostnet [53], and STRING [54], while the number of PPIs would not be increased due to the overlaps and redundancy among the databases.

#### Results and discussion

Figure 3 shows the schematic flow for predicting HPIs beginning with *Homo sapiens* and *Mycobacterium tuberculosis* protein sequences. 138842 pairs of HPIs were

obtained after a BLAST search, then 1863 pairs of HPIs were obtained after DDI filtering, and finally 118 pairs of HPIs were identified after keyword filtering, which involved 43 TB proteins and 48 human proteins.

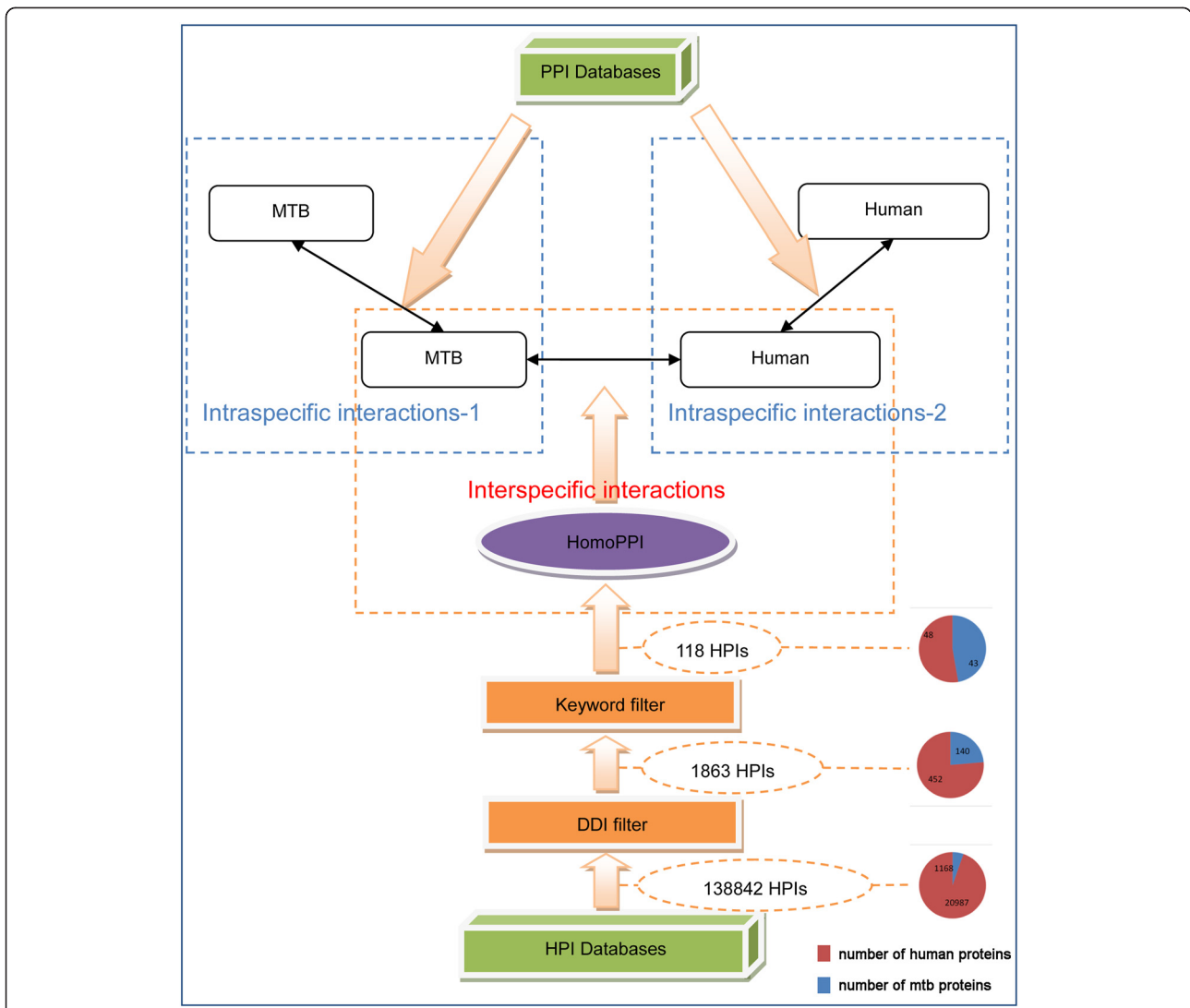
#### The interspecific interactions between *Homo sapiens* and *Mycobacterium tuberculosis*

The HPIs between *Homo sapiens* and *Mycobacterium tuberculosis* (MTB) were predicted based on sequence motifs, using HPIDB and BIPS. By performing a BLAST search of the two databases, 3219 HPIs were obtained between *Homo sapiens* and *Mycobacterium tuberculosis* from HPIDB, and 136664 HPIs from BIPS, with 1041 overlapping HPIs between the two databases. In total, there were 138842 non-redundant HPIs involving 1168 MTB proteins and 20987 human proteins (Figure 4).

Furthermore, the 138842 HPIs were filtered by applying the DDI filter. After aligning the 1168 MTB protein sequences and 20987 human protein sequences to domain or family, 3498 host-pathogen (human-MTB) specific DDIs were extracted (Table 1). Further, by removing redundant HPIs, 1863 non-redundant HPIs were obtained involving 140 MTB proteins and 452 human proteins (Additional file 1: Table S1).

Functional annotations of a protein are important and useful to understand the biological properties. Previous studies indicated that surface proteins consisting of secreted and membrane proteins could play a central role in the interaction of the pathogen with its environment, especially in the pathogenicity of MTB [55], and the term "membrane" was usually used to filter the functional annotation [56-58]. The immune system associated proteins of *Homo sapiens* would also contribute to the host-pathogen interactions [59]. Therefore, functional annotations and biological properties were used to further filter 1863 pairs of predicted HPIs. The "keyword filter" was applied to identify the functional annotation of proteins [60]. The keywords used were "membrane" for filtering *Mycobacterium tuberculosis* proteins, whereas "respiration", "T cell", "lymphocyte", "phagocyte", "lung", "macrophage", "dendritic cell", "immune", "B cell", "alveol", "toll-like receptor", "bronchial epithelial cells" for filtering *Homo sapiens* proteins [48-51]. Each pair of HPI was retained only if both of its interactors corresponded to at least one of the above keywords. Finally, 118 pairs of HPIs were obtained by applying this filtering procedure involving 43 *Mycobacterium tuberculosis* proteins and 48 *Homo sapiens* proteins (Figure 3). All the proteins from *Mycobacterium tuberculosis* engaged in these 118 interactions were associated with the membrane, whereas among the 48 *Homo sapiens* interactors, 8 matched the keyword "T cell", 5 matched by keyword "phagocyte", etc (as in Table 2).

We checked the validity of these predictions by assessing the specificity and sensitivity. Random sets or true



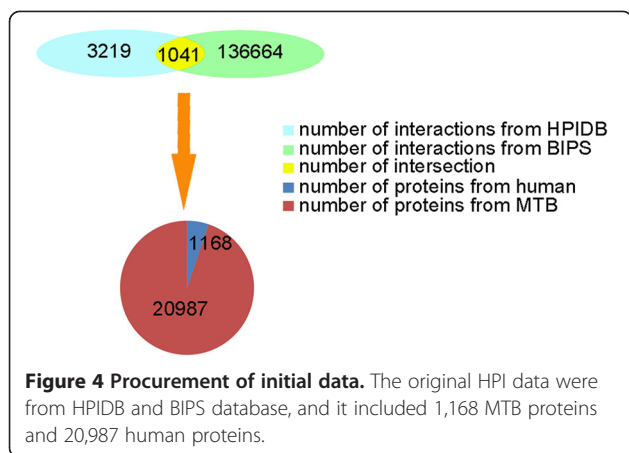
**Figure 3 The systematic flow of HPI prediction.** The Homologous HPI (HomoHPI) were obtained from the HPI databases (HPIDB, BIPS) by BLAST method, followed by applying the DDI filter and keyword filter. After applying these filters, the number of HPIs was trimmed from 138842 to 1863 and then to 118, respectively. The pie charts on the right depict the components of PPIs obtained from each procedure. The intraspecific interactions were extracted from 8 PPI databases (VisANT, Reactome, InteroPorc, IntAct, DIP, MPIDB, MINT, HPRD) and narrowed down to non-redundant data.

negatives were usually used for calculating the specificity [38,61]. In our work, we used the negatome database [62] as a source for non-interactions. 6532 non-interacting pairs from negatome as a reference set were processed by our method including sequence comparison and DDI detection. There were 618 pairs remained after the BLAST step, and they were further narrowed down to 376 pairs after DDI filter. Specificity was calculated as the percentage of correctly predicted true negatives out of 6532 non-interacting pairs. Thus the specificity of our method was 94.2% ((6532-376)/6532). Since gold-standard datasets of experimentally verified human-MTB PPIs are not readily available, we compared our predictions with previous reports to assess the sensitivity and accuracy. Our

predictions included 23 MTB proteins (53.5%) that were suggested to play a significant role in the infection and intracellular survival [50,63-65]. In addition, we also enriched our results with the KEGG pathway and identified more proteins involved in the HPI such as Rv0934, Rv1411c and Rv3875 [66-68]. The coverage of our method depended on the previous experimental observations of similar interactions (template PPI), thus the coverage and accuracy would be increased as more template PPIs were identified.

To improve the accuracy, an increasing number of approaches have been developed taking advantage of the information residing in the motifs or structures. A structure-based interaction network between MTB and human was constructed recently emphasizing the





importance of physical interactions [69]. This structure-based prediction could probably eliminate true negatives, while it was limited by the number of known protein complexes (templates). However, a simultaneous time-course microarray method was developed, which aimed at discovering the HPis experimentally instead of solely depending on the known templates [70,71]. The experiment-based method would make biological sense, while the application of the microarray may not be easy and convenient to any species. All in all, each method would have a good performance in some aspects, and the credibility of known templates was the key point to the “interologs” predictions that mainly based on the sequence comparison.

**The intraspecific interactions among *Homo sapiens* and *Mycobacterium tuberculosis***

For the 43 proteins of *Mycobacterium tuberculosis* and 48 proteins of *Homo sapiens* in the host-pathogen interactions, intraspecific interactions were further studied. The interactions of *Mycobacterium tuberculosis* originated from 7 databases: VisANT, Reactome, InteroPorc, IntAct, DIP, MPIDB, and MINT, whereas the *Homo sapiens* interactions originated from 5 databases: IntAct, HPRD, MINT, Reactome, and DIP. By removing the redundancy from various data sources, there were 587 direct intraspecific interactions in *Mycobacterium tuberculosis* containing 374 MTB proteins and 7157 interactions in *Homo sapiens* containing 3062 human proteins.

**Table 1 The human-MTB specific DDIs in different databases**

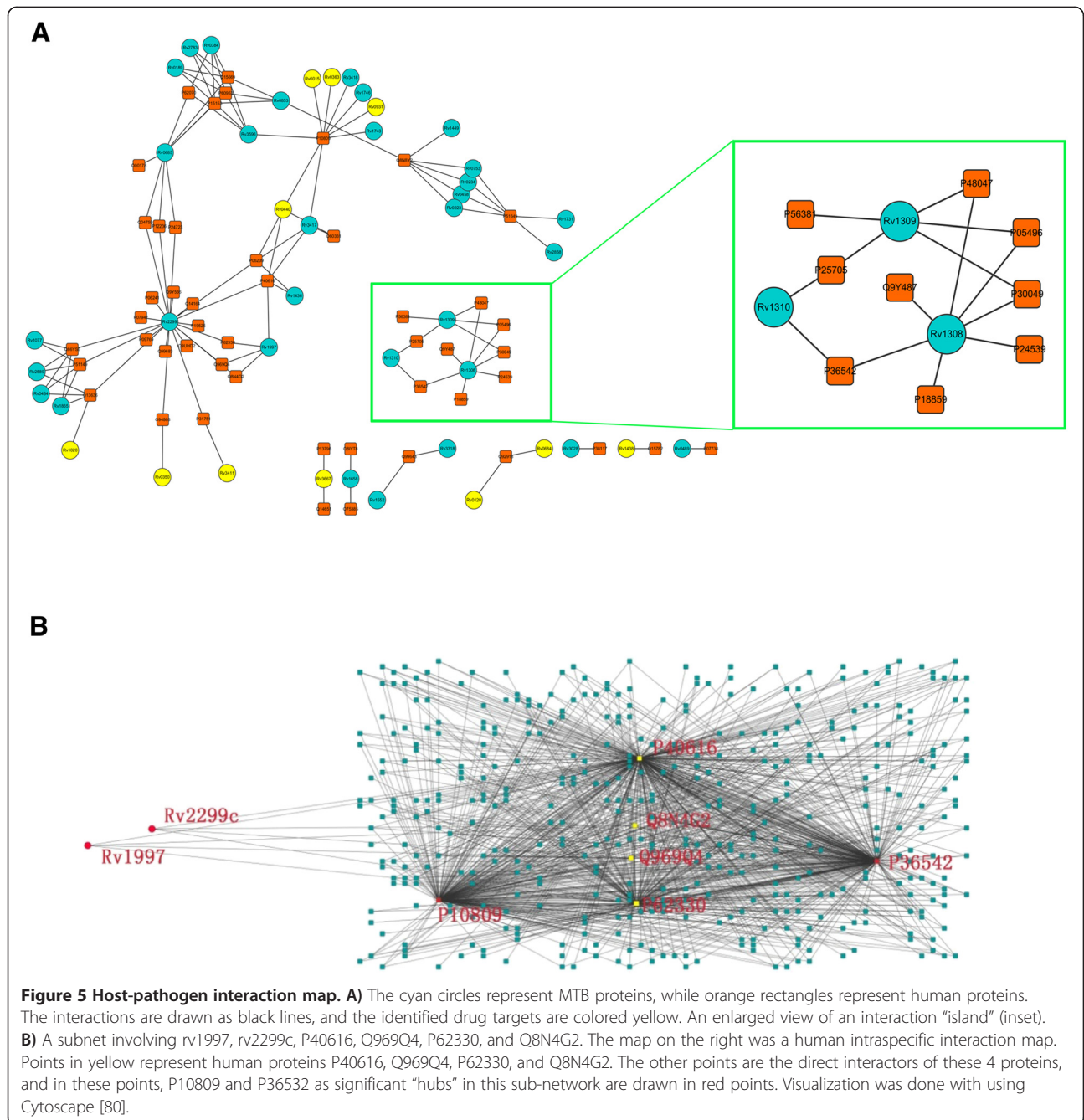
DDI databases	Number of Host-pathogen DDIs
DOMINE	1928
DAPID	62
3DID	601
iPfam	907

**Table 2 The “keywords filter” and its number of corresponding hits**

Keywords	Number of corresponding proteins
Membrane (MTB)	43
Respiration (human)	13
T cell (human)	8
Phagocyte (human)	5
Lung (human)	11
Macrophage (human)	3
Dendritic cell (human)	1
Immune (human)	11
B cell (human)	2
Toll-like receptor (human)	3

**Host-pathogen interaction map and key proteins**

By combining inter-specific interactions with intra-specific interaction, a host-pathogen interaction map was constructed (Figure 5A). MTB proteins rv1308 (atpA), rv1309 (atpG), rv1310 (atpD), which were reported to play significant roles in MTB resistance [72], formed a small “island” in the interaction network by sharing common interactors (Figure 5A), which indicates that these proteins could cooperate with each other to interact with human proteins. MTB protein rv2299c (HtpG), which was predicted to have 20 potential interactors in the network, was previously reported to affect the dormant phase of *M. tuberculosis* [73]. Its interactors, such as P09769, Q14164 and Q9UHD2 protein in human, were identified to be involved in host immune responses based on the functional annotations, which indicated that rv2299c may engage the human immune system. MTB protein rv1997 (ctpF) was detected to be strongly induced during infection of human macrophages [74]. Four interactors (P40616, P62330, Q969Q4, Q8N4G2) of rv1997 (ctpF) mapped in the interaction network were either expressed in the lungs or were involved in immune responses based on the ontology annotation. Figure 5B shows a subset of the interaction map of proteins rv2299c and rv1997, which were also found to share 4 common interactors (P40616, Q969Q4, P62330, and Q8N4G2). These results indicate that proteins rv2299c and rv1997 are essential to understand how MTB survives the host immune response. In addition, human proteins P10809 and P36542 were considered as significant “hubs” and have more interactions in this sub-network. P10809 was previously identified as a key factor, which could influence B cell proliferation, T cell activation and macrophage activation [75-77]. Furthermore, 10 potential drug targets reported before [78,79] were also identified in our network (Figure 5A). It was



noteworthy that 4 MTB targets shared the interactor P10809, which suggested that the human protein P10809 was critical in the MTB infection. Therefore, the predicted HPI map would throw light on how the MTB proteins affect the human cells.

**The structure of PATH**

Although there were many predictions focusing on the HPIs, only a few accessible databases were constructed. To store the predicted host-pathogen-interaction data,

we developed a web-accessible database named PATH (Protein interactions of *M. tuberculosis* and human), which contains not only all the predicted host-pathogen interactions, but also the intraspecific interactions predicted from 7 external databases. Using the web-interface, users can acquire protein-specific interaction information by searching MTB’s gene locus (eg. Rv0001) or Uniprot ID (eg.P49993) and the human protein’s Ensembl identifier (eg.ENSP00000349142) or Uniprot ID (eg.P36542) (Figure 6A). The information of interactors



**Figure 6** Snapshot of the PATH website. **A**) The homepage of the website. Users can acquire interactions information by searching the keywords. **B**) The information of the interactions. It also includes gene ontology annotations from MTB and humans.

both from interspecific network and intraspecific network can also be found during the keyword search (Figure 6B). In addition, the database will be enriched with new HPIs as soon as possible. PATH was built on an Nginx with Python and a MySQL Server as the back-end. HyperText Markup Language (HTML), JQuery and Cascading Style Sheets (CSS) were used at the front-end. It is freely accessible at <http://cadd.pharmacy.nankai.edu.cn/tbdb>. The web server and all parts of the database are hosted at College of Pharmacy, Nankai University, China.

### Conclusion

In this work, we present a specific and integrated database (PATH), which is publicly available and incorporates the predicted interspecific and intraspecific interactions between *Homo sapiens* and *Mycobacterium tuberculosis*. To

our knowledge, PATH is the first specialized database for HPIs on *Mycobacterium tuberculosis*. Our interactions prediction model combined *in silico* algorithms with biological functional annotations. In this study, 118 credible HPIs were identified and stored in the PATH database. In PATH database, users can acquire the interspecific and intraspecific interactions between MTB and human and their related protein interactors by keyword search. The PATH database might facilitate understanding of mechanisms that causes TB, hence help to develop new therapeutic intervention tools for TB.

### Additional file

**Additional file 1: Table S1.** The PPIs derived from DDIs.

**Competing interests**

The authors declare that they have no competing interests.

**Authors' contributions**

TH and JPL made substantial contributions to conception and design, acquisition of data and analysis of data. WL designed the website. JPL and ZHR were involved in drafting the manuscript and revising it critically for important intellectual content and gave final approval of the version to be published. YG and CY agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy and integrity of the work were appropriately investigated and resolved. All authors read and approved the final manuscript.

**Acknowledgements**

We would like to thank staff members of TJAB for their technical support. This work was supported by grants from the State Key Development Program for Basic Research of the Ministry of Science and Technology of China (973 Project Grant Nos 2014CB542800, 2011CB915501 and 2011CB910304).

**Author details**

<sup>1</sup>State Key Laboratory of Medicinal Chemical Biology, Nankai University, Tianjin 300071, China. <sup>2</sup>College of Life Sciences, Nankai University, Tianjin 300071, China. <sup>3</sup>College of Pharmacy, Nankai University, Tianjin 300071, China. <sup>4</sup>Tianjin International Joint Academy of Biotechnology and Medicine, Tianjin 300457, China.

Received: 14 October 2014 Accepted: 13 March 2015

Published online: 26 March 2015

**References**

- Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*. 1998;393(6685):537–44.
- Global tuberculosis report. 2012. [http://apps.who.int/iris/bitstream/10665/75938/1/9789241564502\_eng.pdf].
- Eurosurveillance editorial team. WHO publishes Global tuberculosis report 2013. *Euro Surveill*. 2013;18(43):pii=20615.
- Feltcher ME, Sullivan JT, Braunstein M. Protein export systems of *Mycobacterium tuberculosis*: novel targets for drug development? *Future Microbiol*. 2010;5(10):1581–97.
- Pieters J, Gatfield J. Hijacking the host: survival of pathogenic mycobacteria inside macrophages. *Trends Microbiol*. 2002;10(3):142–6.
- Bodnar KA, Serbina NV, Flynn JL. Fate of *Mycobacterium tuberculosis* within murine dendritic cells. *Infect Immun*. 2001;69(2):800–9.
- Gonzalez-Juarrero M, Orme IM. Characterization of murine lung dendritic cells infected with *Mycobacterium tuberculosis*. *Infect Immun*. 2001;69(2):1127–33.
- Smith I. *Mycobacterium tuberculosis* pathogenesis and molecular determinants of virulence. *Clin Microbiol Rev*. 2003;16(3):463–96.
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001;409(6822):860–921.
- Dyer MD, Neff C, Dufford M, Rivera CG, Shattuck D, Bassaganya-Riera J, et al. The human-bacterial pathogen protein interaction networks of *Bacillus anthracis*, *Francisella tularensis*, and *Yersinia pestis*. *PLoS One*. 2010;5(8):e12089.
- Konig R, Zhou Y, Elleder D, Diamond TL, Bonamy GM, Irelan JT, et al. Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication. *Cell*. 2008;135(1):49–60.
- Bonetta L. Protein-protein interactions: Interactome under construction. *Nature*. 2010;468(7325):851–4.
- Liu X, Liu B, Huang Z, Shi T, Chen Y, Zhang J. SPPS: a sequence-based method for predicting probability of protein-protein interaction partners. *PLoS One*. 2012;7(1):e30938.
- Liu ZP, Chen L. Proteome-wide prediction of protein-protein interactions from high-throughput data. *Protein Cell*. 2012;3(7):508–20.
- Shoemaker BA, Panchenko AR. Deciphering protein-protein interactions. Part II. Computational methods to predict protein and domain interaction partners. *PLoS Comput Biol*. 2007;3(4):e43.
- Lewis AC, Saeed R, Deane CM. Predicting protein-protein interactions in the context of protein evolution. *Mol Biosyst*. 2010;6(1):55–64.
- Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc Natl Acad Sci U S A*. 1999;96(8):4285–8.
- Barker D, Pagel M. Predicting functional gene links from phylogenetic-statistical analyses of whole genomes. *PLoS Comput Biol*. 2005;1(1):e3.
- Galperin MY, Koonin EV. Who's your neighbor? New computational approaches for functional genomics. *Nat Biotechnol*. 2000;18(6):609–13.
- Overbeek R, Fonstein M, D'Souza M, Pusch GD, Maltsev N. The use of gene clusters to infer functional coupling. *Proc Natl Acad Sci U S A*. 1999;96(6):2896–901.
- Aloy P, Ceulemans H, Stark A, Russell RB. The relationship between sequence and interaction divergence in proteins. *J Mol Biol*. 2003;332(5):989–98.
- Liu ZP, Wang J, Qiu YQ, Leung RK, Zhang XS, Tsui SK, et al. Inferring a protein interaction map of *Mycobacterium tuberculosis* based on sequences and interologs. *BMC Bioinformatics*. 2012;13(7):S6.
- Matthews LR, Vaglio P, Reboul J, Ge H, Davis BP, Garrels J, et al. Identification of potential interaction networks using sequence-based searches for conserved protein-protein interactions or "interologs". *Genome Res*. 2001;11(12):2120–6.
- Wang F, Liu M, Song B, Li D, Pei H, Guo Y, et al. Prediction and characterization of protein-protein interaction networks in swine. *Proteome Sci*. 2012;10(1):2.
- Shin CJ, Davis MJ, Ragan MA. Towards the mammalian interactome: Inference of a core mammalian interaction set in mouse. *Proteomics*. 2009;9(23):5256–66.
- Schleker S, Garcia-Garcia J, Klein-Seetharaman J, Oliva B. Prediction and comparison of *Salmonella*-human and *Salmonella*-*Arabidopsis* interactomes. *Chem Biodivers*. 2012;9(5):991–1018.
- Krishnadev O, Srinivasan N. Prediction of protein-protein interactions between human host and a pathogen and its application to three pathogenic bacteria. *Int J Biol Macromol*. 2011;48(4):613–9.
- Hu Z, Hung JH, Wang Y, Chang YC, Huang CL, Huyck M, et al. VisANT 3.5: multi-scale network visualization, analysis and inference based on the gene ontology. *Nucleic Acids Res*. 2009;37(Web Server issue):W115–21.
- Croft D, O'Kelly G, Wu G, Haw R, Gillespie M, Matthews L, et al. Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res*. 2011;39(Database issue):D691–7.
- Michaut M, Kerrien S, Montecchi-Palazzi L, Chauvat F, Cassier-Chauvat C, Aude JC, et al. InterPORA: automated inference of highly conserved protein interaction networks. *Bioinformatics*. 2008;24(14):1625–31.
- Kerrien S, Aranda B, Brezua L, Bridge A, Broackes-Carter F, Chen C, et al. The IntAct molecular interaction database in 2012. *Nucleic Acids Res*. 2012;40(Database issue):D841–6.
- Xenarios I, Salwinski L, Duan XJ, Higney P, Kim SM, Eisenberg D. DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. *Nucleic Acids Res*. 2002;30(1):303–5.
- Goll J, Rajagopala SV, Shiao SC, Wu H, Lamb BT, Uetz P. MPIDB: the microbial protein interaction database. *Bioinformatics*. 2008;24(15):1743–4.
- Ceol A, Chatr-Aryamontri A, Licata L, Peluso D, Briganti L, et al. MINT, the molecular interaction database: 2009 update. *Nucleic Acids Res*. 2010;38(Database issue):D532–9.
- Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, et al. Human protein reference database—2009 update. *Nucleic Acids Res*. 2009;37(Database issue):D767–72.
- Yu H, Luscombe NM, Lu HX, Zhu X, Xia Y, Han JD, et al. Annotation transfer between genomes: protein-protein interologs and protein-DNA regulogs. *Genome Res*. 2004;14(6):1107–18.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215(3):403–10.
- Garcia-Garcia J, Schleker S, Klein-Seetharaman J, Oliva B. BIPS: BIANA Interolog Prediction Server. A tool for protein-protein interaction inference. *Nucleic Acids Res*. 2012;40(Web Server issue):W147–51.
- Kumar R, Nanduri B. HPIDB—a unified resource for host-pathogen interactions. *BMC Bioinformatics*. 2010;11(6):S16.
- Reddy TB, Riley R, Wymore F, Montgomery P, DeCaprio D, Engels R, et al. TB database: an integrated platform for tuberculosis research. *Nucleic Acids Res*. 2009;37(Database issue):D499–508.
- Deng M, Mehta S, Sun F, Chen T. Inferring domain-domain interactions from protein-protein interactions. *Genome Res*. 2002;12(10):1540–8.
- Riley R, Lee C, Sabatti C, Eisenberg D. Inferring protein domain interactions from databases of interacting proteins. *Genome Biol*. 2005;6(10):R89.
- Punta M, Coghill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, et al. The Pfam protein families database. *Nucleic Acids Res*. 2012;40(Database issue):D290–301.



44. Stein A, Ceol A, Aloy P. 3did: identification and classification of domain-based interactions of known three-dimensional structure. *Nucleic Acids Res.* 2011;39(Database issue):D718–23.
45. Finn RD, Marshall M, Bateman A. iPfam: visualization of protein-protein interactions in PDB at domain and amino acid resolutions. *Bioinformatics.* 2005;21(3):410–2.
46. Yellaboina S, Tasneem A, Zaykin DV, Raghavachari B, Jothi R. DOMINE: a comprehensive collection of known and predicted domain-domain interactions. *Nucleic Acids Res.* 2011;39(Database issue):D730–5.
47. Chen YC, Chen HC, Yang JM. DAPID: a 3D-domain annotated protein-protein interaction database. *Genome Inform.* 2006;17(2):206–15.
48. Koo MS, Subbian S, Kaplan G. Strain specific transcriptional response in *Mycobacterium tuberculosis* infected macrophages. *Cell Commun Signal.* 2012;10(1):2.
49. Lee J, Hartman M, Kornfeld H. Macrophage apoptosis in tuberculosis. *Yonsei Med J.* 2009;50(1):1–11.
50. Rohde KH, Abramovitch RB, Russell DG. *Mycobacterium tuberculosis* invasion of macrophages: linking bacterial gene expression to environmental cues. *Cell Host Microbe.* 2007;2(5):352–64.
51. Zuniga J, Torres-Garcia D, Santos-Mendoza T, Rodriguez-Reyna TS, Granados J, Yunis EJ. Cellular and humoral mechanisms involved in the control of tuberculosis. *Clin Dev Immunol.* 2012;2012:193923.
52. Chatr-aryamontri A, Ceol A, Peluso D, Nardozza A, Panni S, Sacco F, et al. VirusMINT: a viral protein interaction database. *Nucleic Acids Res.* 2009;37(Database issue):D669–73.
53. Navratil V, de Chasse B, Meyniel L, Delmotte S, Gautier C, Andre P, et al. VirHostNet: a knowledge base for the management and the analysis of proteome-wide virus-host interaction networks. *Nucleic Acids Res.* 2009;37(Database issue):D661–8.
54. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* 2013;41(Database issue):D808–15.
55. Liu S, Han W, Sun C, Lei L, Feng X, Yan S, et al. Subtractive screening with the *Mycobacterium tuberculosis* surface protein phage display library. *Tuberculosis (Edinb).* 2011;91(6):579–86.
56. Davis FP, Barkan DT, Eswar N, Mckerrow JH, Sali A. Host pathogen protein interactions predicted by comparative modeling. *Protein Sci.* 2007;16(12):2585–96.
57. Dyer MD, Murali TM, Sobral BW. Computational prediction of host-pathogen protein-protein interactions. *Bioinformatics.* 2007;23(13):159–66.
58. Kim JG, Park D, Kim BC, Cho SW, Kim YT, Park YJ, et al. Predicting the interactome of *Xanthomonas oryzae* pathovar *oryzae* for target selection and DB service. *BMC Bioinformatics.* 2008;9:41.
59. Cooper AM. Cell-mediated immune responses in tuberculosis. *Annu Rev Immunol.* 2009;27:393–422.
60. Huo T, Zhang Y, Lin J. Functional annotation from the genome sequence of the giant panda. *Protein Cell.* 2012;3(8):602–8.
61. Simonis N, Rual JF, Lemmens I, Boxus M, Hirozane-Kishikawa T, Gatot JS, et al. Host-pathogen interactome mapping for HTLV-1 and -2 retroviruses. *Retrovirology.* 2012;9:26.
62. Blohm P, Frishman G, Smialowski P, Goebels F, Wachinger B, Ruepp A, et al. Negatome 2.0: a database of non-interacting proteins derived by literature mining, manual annotation and protein structure analysis. *Nucleic Acids Res.* 2014;42(Database issue):D396–400.
63. Rachman H, Strong M, Ulrichs T, Grode L, Schuchhardt J, Mollenkopf H, et al. Unique transcriptome signature of *Mycobacterium tuberculosis* in pulmonary tuberculosis. *Infect Immun.* 2006;74(2):1233–42.
64. Gorna AE, Bowater RP, Dziadek J. DNA repair systems and the pathogenesis of *Mycobacterium tuberculosis*: varying activities at different stages of infection. *Clin Sci (Lond).* 2010;119(5):187–202.
65. Kruh NA, Troutt J, Izzo A, Prenni J, Dobos KM. Portrait of a pathogen: the *Mycobacterium tuberculosis* proteome in vivo. *PLoS One.* 2010;5(11):e13938.
66. Esparza M, Palomares B, Garcia T, Espinosa P, Zenteno E, Mancilla R. PstS-1, the 38-kDa *Mycobacterium tuberculosis* Glycoprotein, is an Adhesin, Which Binds the Macrophage Mannose Receptor and Promotes Phagocytosis. *Scand J Immunol.* 2015;81(1):46–55.
67. Sreejit G, Ahmed A, Parveen N, Jha V, Valluri VL, Ghosh S, et al. The ESAT-6 protein of *Mycobacterium tuberculosis* interacts with beta-2-microglobulin (beta2M) affecting antigen presentation function of macrophage. *PLoS Pathog.* 2014;10(10):e1004446.
68. Ocampo M, Curtidor H, Vanegas M, Patarroyo MA, Patarroyo ME. Specific interaction between *Mycobacterium tuberculosis* lipoprotein-derived peptides and target cells inhibits mycobacterial entry in vitro. *Chem Biol Drug Des.* 2014;84(6):626–41.
69. Ramakrishnan G, Chandra NR, Srinivasan N. From workstations to workbenches: Towards predicting physicochemically viable protein-protein interactions across a host and a pathogen. *IUBMB Life.* 2014;66(11):759–74.
70. Kuo ZY, Chuang YJ, Chao CC, Liu FC, Lan CY, Chen BS. Identification of infection- and defense-related genes via a dynamic host-pathogen interaction network using a *Candida albicans*-zebrafish infection model. *J Innate Immun.* 2013;5(2):137–52.
71. Wang YC, Lin C, Chuang MT, Hsieh WP, Lan CY, Chuang YJ, et al. Interspecies protein-protein interaction network construction for characterization of host-pathogen interactions: a *Candida albicans*-zebrafish interaction study. *BMC Syst Biol.* 2013;7:79.
72. Huitric E, Verhasselt P, Koul A, Andries K, Hoffner S, Andersson DI. Rates and mechanisms of resistance development in *Mycobacterium tuberculosis* to a novel diarylquinoline ATP synthase inhibitor. *Antimicrob Agents Chemother.* 2010;54(3):1022–8.
73. Hegde SR, Rajasingh H, Das C, Mande SS, Mande SC. Understanding communication signals during mycobacterial latency through predicted genome-wide protein interactions and boolean modeling. *PLoS One.* 2012;7(3):e33893.
74. Botella H, Peyron P, Levillain F, Poincloux R, Poquet Y, Brandli I, et al. *Mycobacterial* p(1)-type ATPases mediate resistance to zinc poisoning in human macrophages. *Cell Host Microbe.* 2011;10(3):248–59.
75. Cohen-Sfady M, Nussbaum G, Pevsner-Fischer M, Mor F, Carmi P, Zanin-Zhorov A, et al. Heat shock protein 60 activates B cells via the TLR4-MyD88 pathway. *J Immunol.* 2005;175(6):3594–602.
76. Osterloh A, Meier-Stiegen F, Veit A, Fleischer B, von Bonin A, Breloer M. Lipopolysaccharide-free heat shock protein 60 activates T cells. *J Biol Chem.* 2004;279(46):47906–11.
77. Osterloh A, Kalinke U, Weiss S, Fleischer B, Breloer M. Synergistic and differential modulation of immune responses by Hsp60 and lipopolysaccharide. *J Biol Chem.* 2007;282(7):4669–80.
78. Raman K, Yeturu K, Chandra N. targetTB: a target identification pipeline for *Mycobacterium tuberculosis* through an interactome, reactome and genome-scale structural analysis. *BMC Syst Biol.* 2008;2:109.
79. Vashisht R, Mondal AK, Jain A, Shah A, Vishnoi P, Priyadarshini P, et al. Crowd sourcing a new paradigm for interactome driven drug target identification in *Mycobacterium tuberculosis*. *PLoS One.* 2012;7(7):e39808.
80. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13(11):2498–504.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

