SCIENTIFIC REPORTS

OPEN

SUBJECT AREAS:
SYSTEMS ANALYSIS
DYNAMIC NETWORKS
CELLULAR SIGNALLING NETWORKS

Received
10 April 2013

Accepted
10 July 2013

Published
29 July 2013

Correspondence and requests for materials should be addressed to N.C. (nchandra@biochem.iisc.ernet.in)

Mining large-scale response networks reveals 'topmost activities' in *Mycobacterium tuberculosis* infection

Awanti Sambarey[1], Karyala Prashanthi[2] & Nagasuma Chandra[2]

[1]Molecular Biophysics Unit, Indian Institute of Science, Bangalore -560012, India, [2]Department of Biochemistry, Indian Institute of Science, Bangalore -560012, India.

Mycobacterium tuberculosis **owes its high pathogenic potential to its ability to evade host immune responses and thrive inside the macrophage. The outcome of infection is largely determined by the cellular response comprising a multitude of molecular events. The complexity and inter-relatedness in the processes makes it essential to adopt systems approaches to study them. In this work, we construct a comprehensive network of infection-related processes in a human macrophage comprising 1888 proteins and 14,016 interactions. We then compute response networks based on available gene expression profiles corresponding to states of health, disease and drug treatment. We use a novel formulation for mining response networks that has led to identifying highest activities in the cell. Highest activity paths provide mechanistic insights into pathogenesis and response to treatment. The approach used here serves as a generic framework for mining dynamic changes in genome-scale protein interaction networks.**

Mycobacterium tuberculosis (*M.tb*) kills about 1.5 million people every year and has often been proclaimed as one of the most efficient pathogens known to man[1]. *M.tb* has developed strategies not merely to evade host immune responses but in fact to reside and thrive inside host macrophages. The interactions between the host and pathogen are indeed known to be multi-faceted, involving many feed-back, feed-forward, inhibitory and activatory control structures[2]. The dynamics of interactions between the host and pathogen form a complex web and determine the outcome of an infection. A systems approach is thus essential to comprehend the significance of the multiple events that occur simultaneously, including moves and counter-moves of host and pathogen. The study of host-pathogen interactions has been an area of interest for many pathogens and has been studied through a variety of approaches[3–7].

Several studies of the host transcriptome in response to *M.tb* and other infectious agents are increasingly being reported in literature in recent years[8–12]. Gene expression studies provide lists of differentially expressed genes (DEGs) in a particular condition, but will not provide answers to why or how such an output results. In order to further our understanding and gain mechanistic insights into the causes and consequences of such differential regulation, a systems approach is essential. In a nutshell, the interactions and influences among the various molecular components in a system ought to be captured using a suitable method of abstraction and analysed as a whole[13].

Interaction networks are useful in this regard, and help understand the various connections that exist in a cell. Network modelling has been extensively used in the analysis of biological systems, where the system is mapped on to graphical structures and is a well established field in its own right[14,15]. Networks based on experimentally known protein-protein interactions (PPIs) and theoretically predicted interactions are available for several organisms including different human cell types[16,17]. These networks can provide an overview of the topological architecture in a cell. However, a drawback of this approach is that networks are static snapshots at a given instant of time and they do not readily capture the dynamics that occur in the cell, and hence cannot by themselves explain cellular responses to perturbations such as infections.

The insufficiencies of both approaches can be overcome by integrating genome-wide gene expression profiles into genome scale molecular networks and constructing a specific network for each cellular condition[18]. A protein-protein interaction network by itself is analogous to a street map of a city, while the expression data is analogous to having the types and the numbers of active vehicles in the city at a given time point. The two types of data are clearly complementary to each other and provide different perspectives of essentially the same system. If

integrated together by superimposing gene expression or proteomic data onto topological networks, the insights they can provide will be significantly greater than what either of them can provide individually.

Here, we seek to model the flow of information in response to tuberculosis (TB) infection. Relating to the same analogy, the objective of this work would be comparable to looking for 'highest vehicular activities' in a city's roads upon a particular trigger. We apply graph theoretical methods to reconstruct weighted directed networks by integrating 'omics' data from tuberculosis patients available in literature with a curated large scale model and identify 'highest activity' paths in the cell.

Results

We first reconstructed an elaborate model of the macrophage upon exposure to *M.tb* to capture important functional processes and their interconnections including pathogen uptake, consequent signalling and activation of the appropriate immune response. Macrophages are constituents of peripheral blood mononuclear cells (PBMCs), which also include dendritic cells and lymphocytes. This constructed generic model of the macrophage is largely applicable to PBMCs as well. Gene expression data from clinical TB patient samples was then taken from literature to construct specific response networks. Dynamic changes occurring upon infection and over the course of treatment with drugs were captured and compared with appropriate healthy controls to identify routes or paths with highest activities in these conditions.

A general extensive molecular systems model of the macrophage. Contact of *M.tb* with the host occurs in the lungs, and alveolar macrophages along with neutrophils and lung dendritic cells (DCs) are the first host cells to interact with *M.tb*. They secrete several cytokines, chemokines and produce antibacterial peptides upon infection. These initial innate immune responders subsequently govern the activation of T cells, components of the adaptive immune response. Under the influence of the cytokine IL12 and chemokines CCL19 and CCL21, *M.tb* infected DCs migrate to lung draining lymph nodes and drive the differentiation of naive T cells to the T_h1 phenotype[19]. After approximately two weeks of initial infection, antigen-specific T_h1 cells migrate back to the lungs under chemokine influence and mount a protective response by producing IFN-γ[20], resulting in macrophage activation leading to augmented cytokine and chemokine production, release of reactive oxygen intermediates (ROIs) and reactive nitrogen intermediates (RNIs), as well as other effector molecules[21–23]. The T cell dependent response comprises CD4+ and CD8+ T cells which recognise antigenic peptides on MHC class II and MHC class I molecules respectively, as well as CD1 restricted T cells that recognise glycolipids present in mycobacterial cell walls. $\gamma\delta$ T cells, an independent lymphocyte population, are CD3+ T cells which recognise antigens independently of specific presentation molecules. Along with DCs, $\gamma\delta$ T cells participate in the early immune response to *M.tb* and produce pro-inflammatory cytokines such as IFN-γ and TNF-α[24]. Thus, the components of the innate immune system are key elements in initiating and influencing the ensuing adaptive response that mediates protective immunity against *M.tb* infection[25].

We constructed a network model of the processes triggered in the macrophage in response to tubercular infection, based on an extensive literature survey. The model captures responses in the macrophage, commencing with phagocytosis involving the uptake of *M.tb* or mycobacterial components mediated by different surface receptors on the macrophage for both opsonised and non-opsonised bacteria. Some of the receptors include complement receptors (for e.g. CR1, CR3), mannose receptors (MR), type A scavenger receptors (e.g. SR-A1, SR-A2) and surfactant-protein A (SP-A)[26]. The immune response put forward by the phagocyte is dictated by the receptor involved in *M.tb* uptake. Uptake of *M.tb* by CR1 has been shown to provide a better chance for pathogen survival than by CR3 or CR4[27]. Mycobacterial uptake in the presence of surfactant-protein A or through mannose receptors renders the bacteria better poised to repress reactive nitrogen intermediates and reactive oxygen species respectively, both of which are major methods of mycobacterial elimination by the macrophage[27]. The nature of interactions of the receptors with *M.tb* is thus complex and dynamic, not only being responsible for uptake but also for triggering or inhibiting cellular signalling networks that mediate processes of bacterial elimination. The molecular details of these processes are derived from several experiments that typically use non phagocytic cells transfected with the receptors of interest[28,29]. In the real scenario, however, the uptake by phagocytes and the downstream events depends upon a number of aspects including ratios of receptor abundances available on the phagocyte cell surface, and their relative strengths to recognise pathogenic molecules, thus making it important to study the system as a whole.

Different signalling mechanisms such as calcium and phosphoinositide signalling are activated upon binding of *M.tb* to the macrophage, which regulate cytoskeleton rearrangement as well as vesicular trafficking resulting in the formation of the phagosome[30,31]. The phagosome then acquires the early and late endosome markers, followed by fusion with the lysosome to form the phagolysosome[32]. Fusion with the lysosome leads to bacterial death, and subsequent processing and presentation of antigens through the MHC presentation pathway[23].The antigens presented are recognised by T cell receptors, co-stimulatory molecules (CD80/CD86) and adhesion molecules (intracellular adhesion molecule ICAM-1)[23]. T cell activation is central to the protective immunity to *M.tb* infection, and cytokines play an important role in driving T cell differentiation[33]. Interleukin-12 (IL-12) stimulates the formation of T_h1 cells whereas interleukin-4 (IL-4) triggers the formation of T_h2 cells[34,35]. These T_h cells release cytokines having varying effects: the T_h1 related cytokines (for e.g. IFN-γ, TNF-α, IL-2) are pro-inflammatory cytokines (PICs), while T_h2 related cytokines (for e.g. IL-4, IL-5, IL-10, IL-13) are generally anti-inflammatory in nature (anti-inflammatory cytokines or AICs)[33–35]. Acute cytokine responses ensue upon *M.tb* infection of macrophages, triggered through phagocytic and non-phagocytic interactions and *M.tb* components. As the immune response develops over time, the cytokine and chemokine profiles show variation, and their activity may be beneficial or detrimental depending on their relative times of production[36]. Cytokines and chemokines may act synergistically or antagonistically. Generally, Th1 related cytokines mediate protective response to *M.tb*, while Th2 responses regulate and inhibit Th1 responses. T_h2 related cytokines can also inhibit the production of chemokines produced by the macrophages[37,38]. Chemokines are major signalling molecules responsible for the progress of inflammation, regulating lymphocyte infiltration and stimulating trans-endothelial migration of T lymphocytes[39]. The interplay between PICs, AICs, chemokines and regulatory cytokines is thus significant in governing the outcome of disease.

Many mycobacterial components trigger or inhibit the host response towards elimination of *M.tb*, enabling bacterial survival. Mycobacterial cell wall components Lipoarabinomannan (LAM), Mannosylated LAMs (ManLAM), Phosphatidylinositol Mannoside (PIM) and Trehalose 6,6′-dimycolate (TDM) are the major modulators of phagosome maturation and other signalling pathways[40–42]. ManLAM prevents fusion of mycobacterial phagosome with the late endosome and lysosome by inhibiting the Calmodulin-Ca2+ phosphotidol-inositol-3 kinase[43]. PIM promotes the interaction and fusion of mycobacterial phagosome with early endosome, so as to deliver nutrients to mycobacterial phagosome and maintain the non-acidic pH[44]. TDM also inhibits phospholipid vesicle fusion[45]. ManLAM activates host p38MAPK and PPAR signalling pathways that activate prostaglandin E2 (PGE2), bringing about inhibition of

Th1 response, ROI, NF-κB, and activation of Th2 response leading to the formation of alternatively activated macrophages[46]. LAM activates TLR2/1 and brings about the expression of PICs via the transcription factor NF-κB[42]. LAM is also known to activate ERK1/2 MAPK pathway thereby leading to activation of ROI and RNI[42]. TLR activated NF-κB mediates synthesis of WNT5A which interacts with FZD5, regulating the microbial-induced IL-12 production in a paracrine manner, thereby activating T cells that are primed specifically for IFN-gamma release[47,48]. *M.tb* also regulates host antigen presentation[49].

The immune response to *M.tb* is complex and involves various host and bacterial factors, as well as multiple signalling pathways and heavy cross-talk between them. While the model presented here describes the general processes triggered in response to *M.tb* infection, it is still an oversimplified depiction of the components and their complex interactions in the macrophage. Nevertheless, the model captures the essence of biological processes important for host-pathogen interactions and presents a framework to study it from a systems perspective. The model is amenable to expansion and inclusion of transcriptomic data as it becomes available.

Figure 1 (a) provides a schematic representation of the model comprising the biological processes or modules triggered in response to *M.tb* infection as described, and these host-pathogen interactions collectively dictate the consequences of infection. The model constitutes 1888 proteins from 32 modules that capture the different cellular signalling processes described above. The distribution of proteins in the 32 modules is shown in Figure 1 (b). We then derived a network of interactions of these proteins based on experimental evidence from literature and several protein-protein interaction databases. Directionality was assigned to the edges wherever applicable based on known experimental information. The final curated network comprises 1888 nodes (proteins) and 14016 edges (protein-protein interactions). Of these, 11,283 edges were assigned direction, while the remaining 2733 interactions were kept as bidirectional (The final list of interactions along with information of their sources and direction are given in the supplementary table S1).

Topological network analysis indicated that the network was scale-free, allowing the identification of highly connected 'hub' nodes[50]. The edge-betweenness, an important measure of centrality for edges in the network was computed and considered for further analysis (Network properties are given in Supplementary table S3).

Response networks indicate differential enrichment of modules. Recent genome-wide transcriptional studies have contributed to our current understanding of tuberculosis-specific responses in the host. To this end, a study by Berry et al. (2010)[8] reported significant differences in gene expression values of several genes across different conditions of disease and treatment. The authors identified a signature of 86 differentially expressed genes unique to tuberculosis infection. More importantly, a clustering analysis of the profiles was able to group healthy samples separately from the disease samples, while also illustrating that the expression profile of the 12 month drug treated samples was closer to that of healthy controls than active disease. It is thus clear that the observed gene expression changes are specific and characteristic of the condition. Microarray analysis however, is insufficient in providing explanations for the cause or consequence of differential expression in the identified gene set.

To overcome this limitation, gene expression changes reported by Berry et al. (2010) for 6 different conditions were integrated into the directed network to generate response networks. These were constructed first by incorporating gene expression levels for each condition as node weights, followed by computation of edge weights as a function of both the node weight of connected nodes and the corresponding edge betweenness. We define six weighted response networks, each incorporating gene expression data from (a) monocytes from healthy controls (HC_M), (b) monocytes from pulmonary tuberculosis patients (PTB_M) (c) whole blood from healthy controls (HC_B) and whole blood from tuberculosis patients monitored through drug treatment over (d) 0 months (PTB_0), (e) 2 months (PTB_2) and (f) 12 months (PTB_12).

Advances in microarray technology and standardisation procedures have made it feasible to compare signal intensities across conditions of similar samples for a given platform and chip[51]. Hence for this study, the normalised signal intensity for each gene, together with its comparative fold change were considered as opposed to fold change alone, in order to capture relative expression levels of different genes in a given condition, as well as their extent of change between conditions. Constitutively expressed genes that do not vary significantly in different conditions are also automatically captured with this scheme. The edge weight is thus representative of both expression values and network topology. In our scoring scheme, a low edge weight corresponds to high expression levels of both nodes contributing to the edge, as well as high edge centrality, thereby indicating a higher functional importance of the edge. (Node and edge weights for the response networks have been provided in supplementary tables S2 and S1 respectively.)

The significance of changes in expression for individual nodes in the network can be inferred better when viewed in the context of their functional modules. Module-wise networks were constructed by computing cumulative weights of individual nodes within the modules. Figure 2 illustrates the effects of weighting module-wise networks in different conditions. The module-module influence network for the condition PTB_M, as shown in Figure 2 (a), illustrates the extent of modular expression for the entire network which is the cumulative node weight of all nodes in each module, while the edge thickness represents the extent of interaction between two modules. Figure 2 (b) and (c) provide a comparison between module expressions in monocytes and in blood from patients over treatment respectively. The figure illustrates that modules Phagocytosis, Cytokine signalling, JAK-STAT signalling, TLR signalling, ROI and SNARE have significantly higher weights in disease as compared to healthy controls and treated samples. Antigen presentation is seen to have a significantly lower weight in PTB_M, owing to the drastic difference in the node weight of HLA-DRB5, with over 500 fold decrease in expression seen in PTB_M in comparison to HC_M (Supplementary table S2). A full list of differences in modular expressions for different module-module influence networks is given in Supplementary tables S4 and S5.

Network mining reveals 'Top most activities' in the cell. In order to gain insights into possible functional roles of the observed differential regulation, we implemented graph theoretical approaches to identify connected paths of highest activities anywhere in the cell. Identification of such paths was posed as a problem of computing shortest paths in each of the weighted response networks. Shortest paths among all nodes in the networks were computed using Dijkstra's algorithm as described in the methods section. The algorithm finds the shortest path from a given source node to a target node in the network by traversing edges with the lowest edge weight. Depending on the edge weights, which are reflective of the expression levels as well as centrality of the edge, different routes may be taken in different response networks. We introduce the term *Path Cost* which is the cumulative sum of the weights of the edges constituting the path, described in equation (7). Lower the *Path Cost*, higher is the perceived activity through that path.

Among all pair shortest paths which amount to 34,13,442 paths in each individual response network, we considered paths of length 2 and above. We rank listed and chose top most paths by applying a threshold derived by analysing the distributions of *Path Costs* of all shortest paths (Figure 3(a)). Three cost ranges are visible in the figure: (a) less than1, (b) costs between 1 and 2 and (c) those with costs greater than 2. In the PTB_M network, 5862 paths appeared to

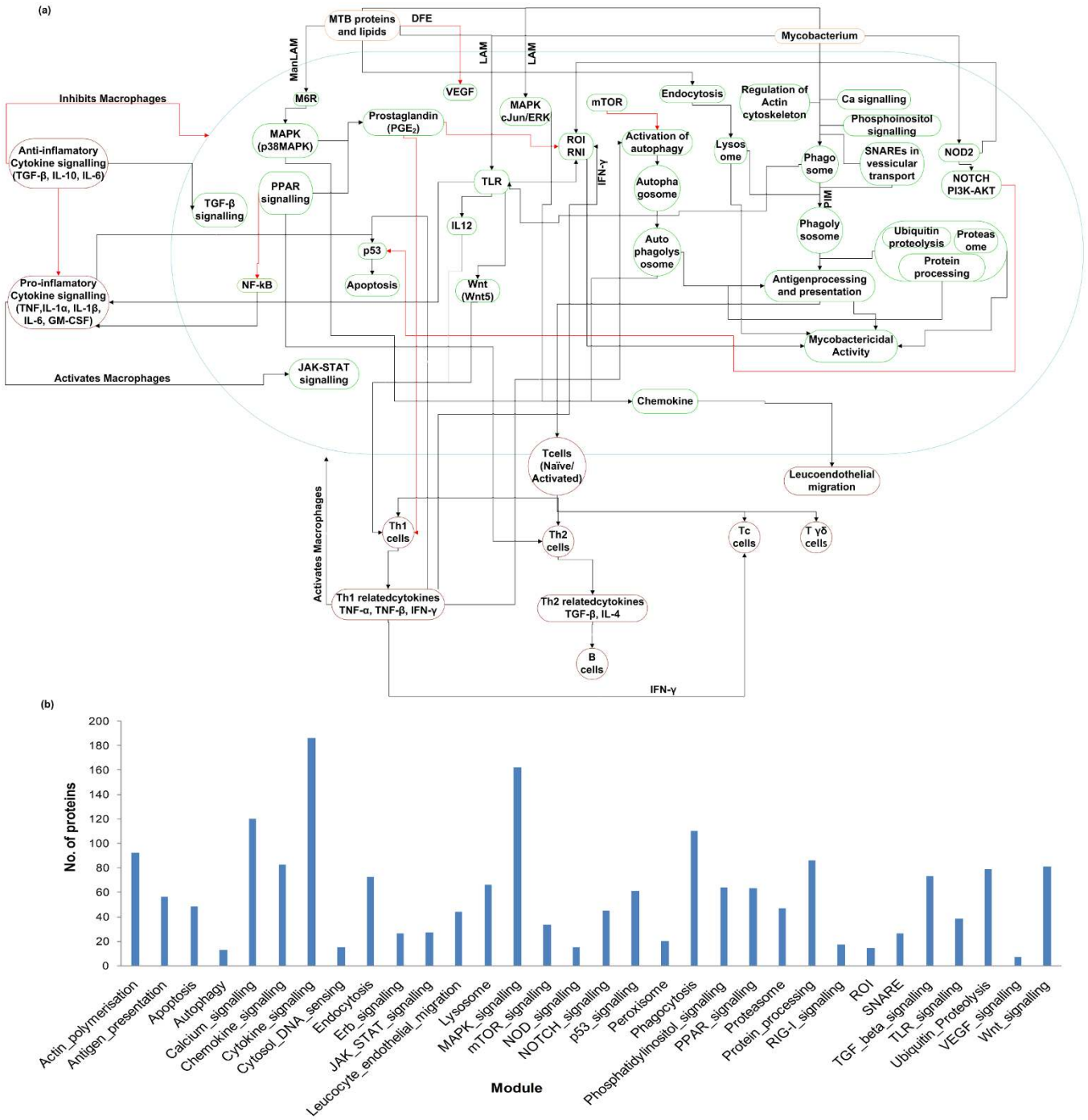


Figure 1 | Modelling the macrophage response in tuberculosis. (a) A molecular map describing responses triggered in the macrophage upon *M.tb* infection. A single macrophage is reflected in green, and elicits both intracellular (outlined in green) and extracellular (outlined in purple) responses upon encountering either complete mycobacteria or mycobacterial lipids and other components (highlighted in orange). Red arrows represent inhibitory influences, while black arrows indicate activation. (b) Distribution of 1888 proteins in 32 biological modules participating in the cellular response described in (a). Here, the module Cytokine shows highest representation of proteins while VEGF signalling has the least representation, in terms of number of proteins involved in that process.

be of least cost (<1), amounting to 0.17% of all shortest paths in the network, and are tagged as highest activity paths hereafter. Those with *Path Costs* between 1 and 2 made up a further 0.87% and reflect reasonably high activity paths, while those >2 are considered to be relatively lower activity paths. Thus, the thresholds were chosen based on the observed distribution but are still to an extent arbitrary, erring on the side of caution so as to ensure that the selected paths are of highest activity, while possibly allowing some high activity paths but with higher *Path Costs* to be missed out. High activity paths for all conditions are provided in Supplementary tables S6–S11.

Module representations in the highest activity paths. In order to observe module-wise activities in the highest activity paths, percentage distribution of nodes belonging to different modules in paths below a cost of 1 was calculated. Figure 4 shows the percentage representation of modules in all paths below 1 for different

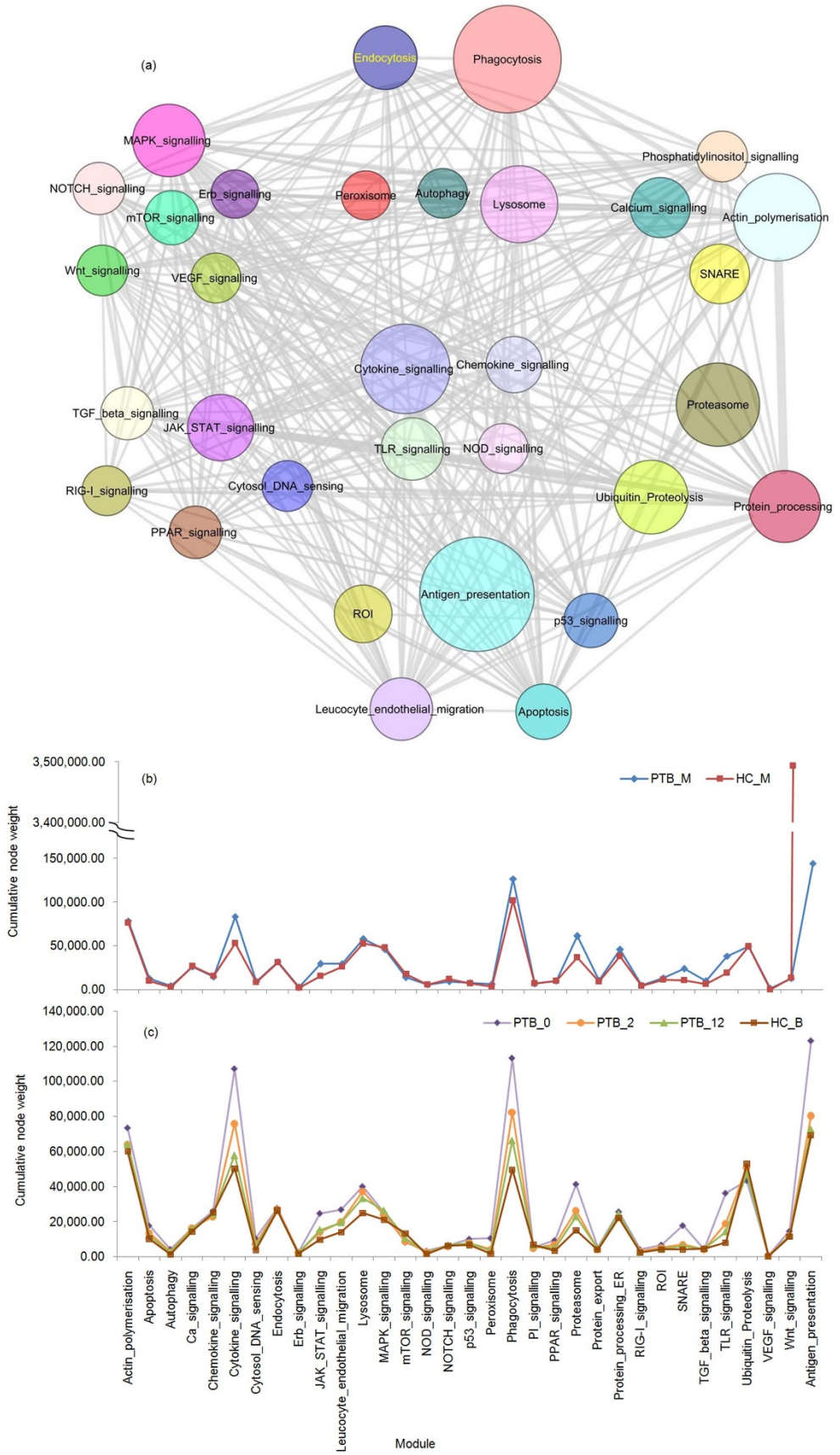
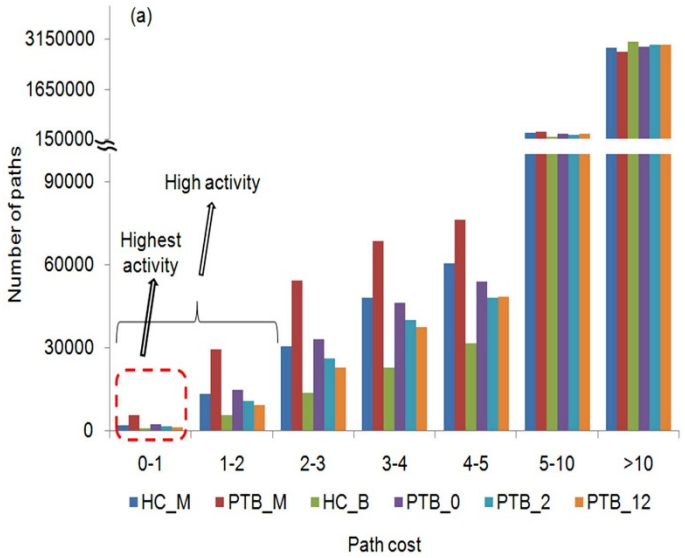


Figure 2 | Response networks and inter-module influences. (a) Module-module influence network for PTB_M. Node sizes are reflective of the number of nodes and their cumulative expression in each module, while edge thickness indicates the number of influences between the connected modules and their interaction strength. Comparative expressions of 32 modules in (b) PTB_M and HC_M and (c) different states of treatment viz. PTB_0, PTB_2 and PTB_12 and healthy controls HC_B. In Figures 5, 6 and 7, each module is coloured differently indicated by the module labels in Figure 2(a).



(b)

Condition	Path Cost	
	0-1	1-2
HC_M	2243	13408
PTB_M	5862	29697
HC_B	967	5750
PTB_0	2551	14990
PTB_2	1825	11031
PTB_12	1463	9272

Figure 3 | Path Cost distribution reveals differences in activities. (a) *Path cost* distribution across all 6 conditions. Paths with least cost, in the range 0–1 are considered to be of 'highest activity', while those in the range 1–2 are relatively 'high activity' paths. (b) Table showing number of paths falling in the highest and high activity regions respectively, for all conditions.

conditions. The figure provides a ready comparison of the various changes that occur in 'highest activity' paths in monocytes from tuberculosis patients (PTB_M) as compared to those from healthy volunteers, as well as those activities that vary across treatment. Several changes in PTB samples are evident from this analysis, consistent in the high activity paths of both PTB_M and PTB_0 such as: (a) significant increase in activities of modules Cytokine_signalling, JAK_STAT_signalling, Lysosome, TLR_signalling, Proteasome,

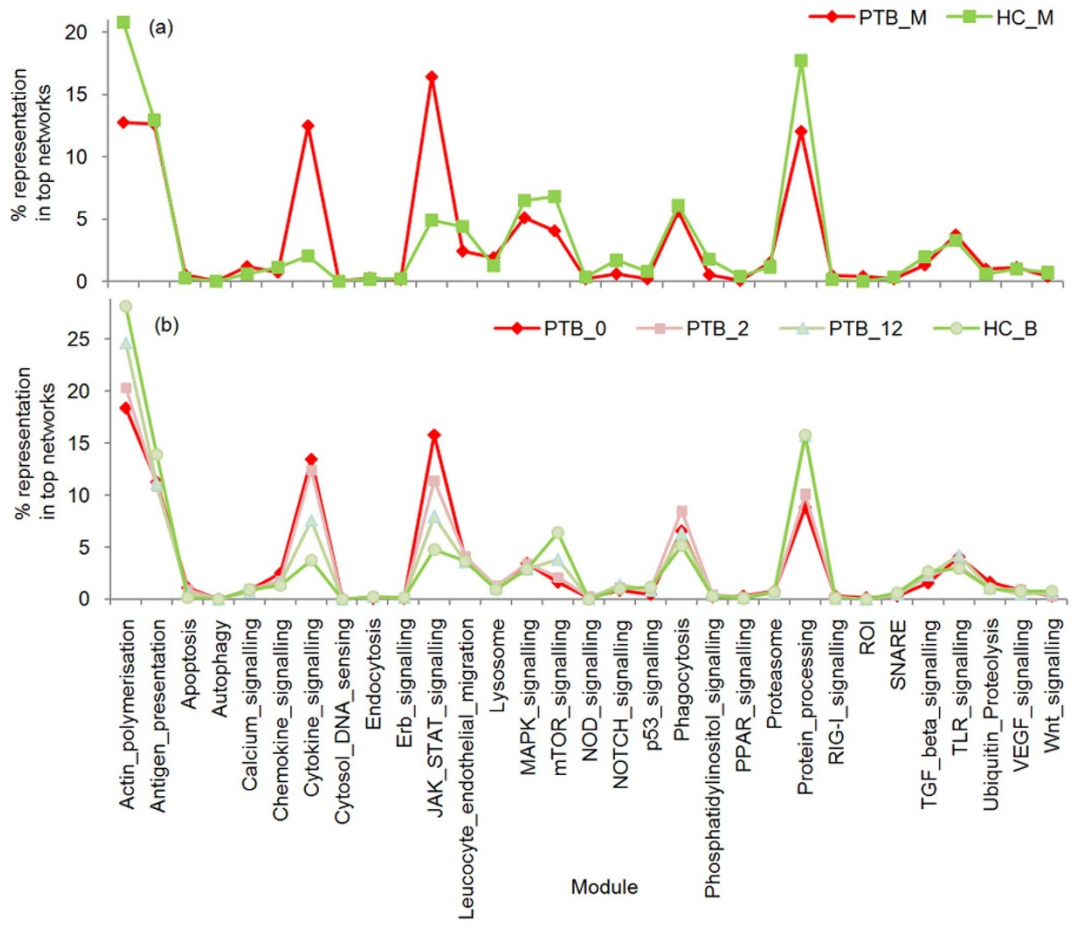


Figure 4 | Module representation in highest activity paths. (a) Percentage representation of modules in all the highest activity (*Path Costs* below 1) paths shown for monocytes in healthy conditions as well as in disease. (b) Percentage module representation in the highest activity paths upon treatment and in healthy controls.

Ubiquitin_Proteolysis, RIG-I_signalling, VEGF_signalling, Cytosol_DNA_sensing and Apoptosis (b) consistent decrease in Actin_polymerisation, Antigen_presentation, Erb_signalling, MAPK_signalling, mTOR_signalling, SNARE, p53_signalling and Protein_processing. In PTB_M alone, we see an increased representation of the modules Endocytosis, Calcium_signalling, ROI and Autophagy as compared to their healthy counterparts, while in PTB_0 we see a higher representation of Chemokine_signalling, NOD_signalling and PPAR_signalling, with their activities decreasing over treatment. PPAR signalling has been shown to be exploited by *M.tb* to enable survival and lead to active disease[46]. Some modules such as Wnt_signalling do not show much variation in disease and healthy conditions. Module representations in the top paths thus substantially concur with known response profiles. Several modules are seen to revert their activities upon treatment , similar to the activities seen in HC_B top paths, such as JAK_STAT signalling, Cytokine signalling, mTOR signalling and Actin polymerisation (Figure 4b), indicative of a disease specific response which highlights the most regulated biological modules in tuberculosis infection.

Comparison of these module activities with their node weights in Figure 2(b) indicates that high module expression alone does not necessarily correlate with increased modular activity in the top paths, as can be seen in the resulting decrease in SNARE activity in the PTB_M and PTB_0 top paths, despite having a higher modular node weight in both, highlighting the importance of a network perspective in interpreting large-scale expression data.

Broadly, all of these observations from the response networks are consistent with knowledge from literature based on a vast number of individual experimental studies[25] and can thus be rationalised. Module-wise analysis provides an overview of the processes occurring within the cell while the actual paths provide insights into which specific molecular events could be occurring.

Top paths themselves form connected networks. A network was constructed using only the highest activity paths for PTB_M. 325 nodes and 501 edges from 5862 paths were seen to comprise the PTB_M high activity network, shown in Figure 5(a) while 268 nodes and 381 edges from 2243 paths constituted the highest activity network in HC_M (Figure 5(b)). A small fraction of the paths (23.1%) remain common to both PTB_M and HC_M highest activity networks, and a sub-network of paths uniquely present in the PTB_M top network was constructed, representing paths that are specifically activated in PTB_M. Figure 5(c) shows the sub-network of highest activity paths unique to PTB_M, constituting 208 nodes and 286 edges.

It is intriguing to observe that the highest activity paths form well connected networks by themselves. It must be noted that only 11% of nodes from the complete network are involved in the uniquely active network for PTB_M. Yet they remain well connected, implying a smooth flow of information through these paths. The connectedness provides a basis to understand the purpose (or consequence) of up-regulation of a given node and how the perceived increase in activity is transmitted to its neighbouring nodes forming a highly active path. The PTB_M top path network is denser than the HC_M top path network, indicating an increased upsurge in activity in disease, and leads to the identification of participating nodes in these active paths. It is also readily apparent from these networks that the overall extent of activity is significantly higher in PTB than in a healthy state.

Fold change in gene expression is insufficient to convey extent of participation in a path. On comparison of the HC_M and PTB_M highest activity networks, 97 nodes were seen to be unique to the PTB_M highest activity network, while 40 nodes were unique to the HC_M top network. Interestingly, 228 nodes were seen to be commonly occurring in both PTB_M and HC_M top networks. Nodes occurring uniquely in the PTB_M top network showed a range of variation in gene expression as compared to their healthy counterparts. Of the 97 unique nodes, 62 nodes were up-regulated (Fold change > 2) in PTB_M with high signal intensities, while the remaining 35 nodes did not show significant variation in gene expression between HC_M and PTB_M. Since the node weights have been derived by considering normalised signal intensities in addition to fold change values, they are able to reflect approximate abundances of individual proteins. We must of course note that gene expression values cannot directly be translated into protein concentrations, but for lack of any more comprehensive quantitative data for studies of this scale at present, we assume that these expression values are reflective of protein activity in the cells.

Upon a thorough examination, it becomes apparent that the fold change values do not by themselves reflect extent of function mediated by that node. While the appearance of highly up-regulated nodes in the top-activity networks can be justified, the top networks of both HC_M and PTB_M see nodes that are constitutively expressed in these conditions. In fact, 228 nodes are common to both these networks, which may seem surprising at first and would have been missed out completely by considering only the fold change values. However, it becomes clear that these nodes, many of them being constitutively expressed genes, form high activity bridges between up-regulated nodes forming top scoring paths. Thus, considering DEGs alone would not be sufficient to get a network perspective of activities in a cell (the complete list of nodes in the top networks for PTB_M and HC_M, along with their weights is given in Supplementary tables S12–S14).

Nodes occurring consistently in tuberculosis networks. In order to identify nodes appearing consistently in networks for disease, we inspected the top-path networks over treatment. Two conditions were considered – a) Nodes that occur in the PTB_0 top path network and also occur uniquely in the PTB_M highest activity network, but not in the HC_M, HC_B or PTB_12 highest activity networks. b) Of the nodes shortlisted in a), those that show maximal difference in their node weights in between PTB_0 and PTB_12 were identified, suggesting that these nodes respond to treatment. By considering the top path network for both PTB_M and PTB_0, we extracted those nodes in the host that are significant, consistent and specific contributors to the host's tuberculosis response. Table 1 provides a list of these nodes showing significant differences in node weights upon treatment, along with the corresponding module that the node belongs to. Thus, a testable hypothesis is obtained from this study in the form of a shortlist to be explored as possible markers of active tuberculosis.

Network re-rerouting upon treatment- 'Part Reversal' to non-infected state. One benefit of analyzing response networks in the form of active modules is that it enables monitoring the modular activity upon treatment. Networks corresponding to response to treatment at 0, 2 and 12 months are constructed by incorporating appropriate node weights obtained from gene expression profiles of the clinical samples as reported by Berry et al. (2010). Significant differences are indeed seen in the path activity profiles in treatment related response networks PTB_0, PTB_2 and PTB_12. Overall, we observe that module profiles and paths that occur in the PTB_12 networks are more similar to those in healthy controls (HC_B) (Figure 4(b)). Many paths that were seen to emerge in PTB_0 due to low *Path Costs* have reversed back to high costs upon treatment, especially with PTB_12.

Some examples of nodes in such paths along with their immediate neighbourhoods in different conditions are illustrated as sub-networks in Figure 6. Figure 6 shows an example sub-network for the activation of IFN-γ target genes PIM1, CD40 and GBP1 by IFN-γ receptors (1 and 2). IFNγ-R2 appears to be more active than IFNγ-R1 in both PTB_M and PTB_0 and activates the target genes through the JAK-STAT signalling pathway. However, in the corresponding healthy controls, as there is no infection the IFN-γ receptors and

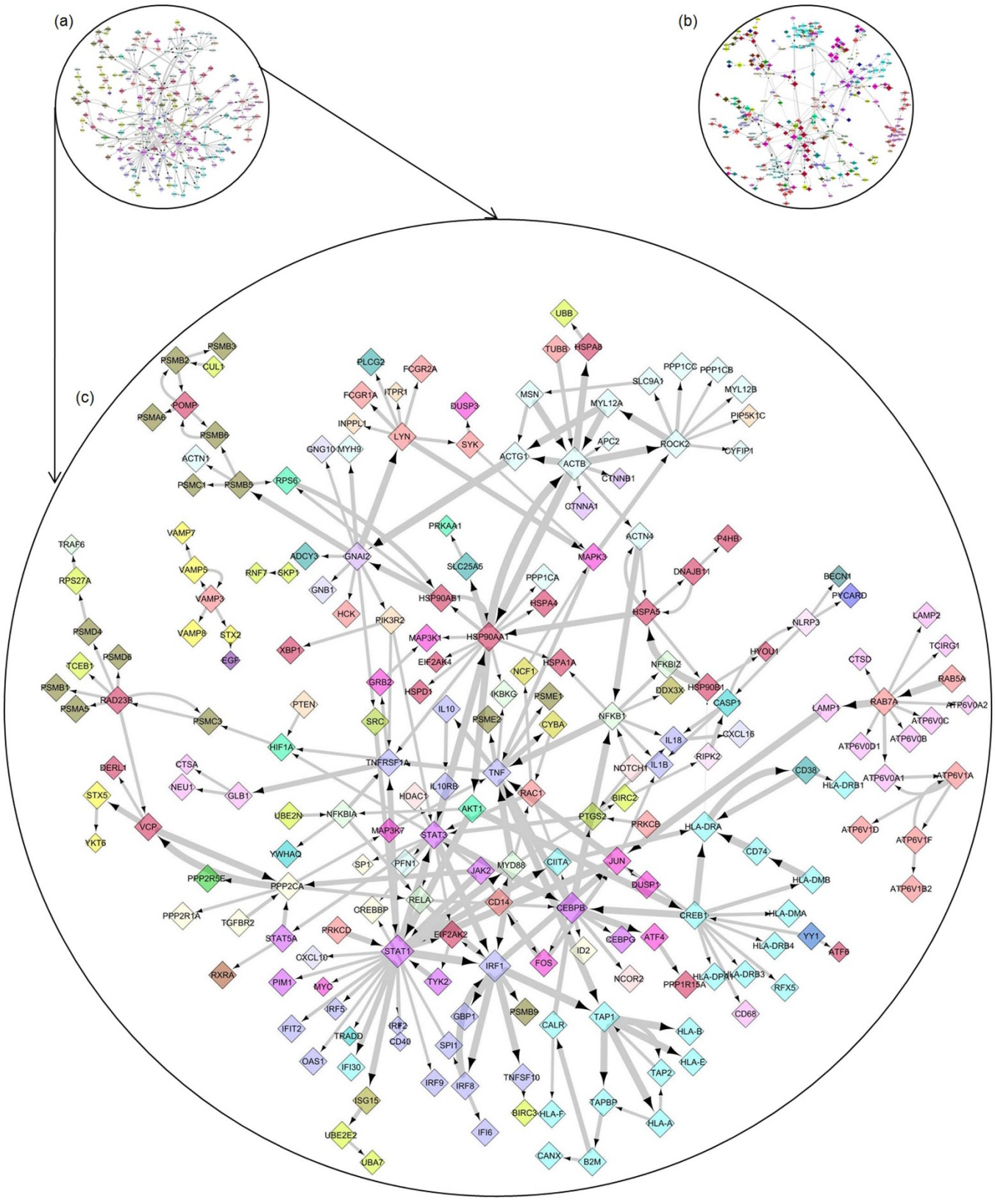


Figure 5 | Highest activity paths in PTB_M and HC_M forming well connected subnets. Network of nodes and edges constituting highest activity paths with *Path Costs* below 1 for (a) PTB_M and (b) HC_M. (c) A zoomed-in sub-network representing unique highest activity paths in PTB_M. Nodes are coloured according to modules they belong to, and the edge thickness reflects the strength of the edge.

Table 1 | A short list of characteristic nodes seen in tuberculosis top networks. Nodes occurring in PTB_0 and PTB_M highest activity networks, but not in HC_M, HC_B, or PTB_12 highest activity networks that have the potential to serve as possible biomarkers for disease. Nodes showing maximal difference in node weights upon treatment are represented here

Node	Node weight PTB_0	Node weight PTB_12	Module
FCGR1A	8912.80	173.81	Phagocytosis
VAMP5	10789.79	446.83	SNARE
CASP5	877.60	56.98	Apoptosis
OAS1	680.09	50.63	Cytokine_signalling
GNG10	2343.51	243.23	Chemokine_signalling
NCF1	4405.63	689.49	ROI
IRF7	1659.32	317.85	Cytokine_signalling
IFI30	2016.60	437.90	Antigen_presentation
IRF2	144.59	47.95	Cytokine_signalling
PYCARD	2029.13	681.97	Cytosol_DNA_sensing
DUSP3	1809.13	667.71	MAPK_signalling
CCR1	777.42	323.87	Chemokine_signalling
VAMP3	1484.29	718.59	SNARE
HLA-F	3684.13	2056.94	Antigen_presentation
IRF9	3197.67	1857.75	Cytokine_signalling

the target genes have very low node weights and higher edge weights, so the algorithm takes a different, longer route through low-cost edges while computing the optimal shortest path to the given target genes, which may not be functionally significant as seen in paths leading to CD40. This can be explained better in the context of treatment, with the activity of IFN mediated responses subsiding over the course of treatment, as the state of infection is lost, approaching a profile similar to the healthy controls. A similar scenario for differential path activity for downstream signalling through IFNA receptors is shown. The pathways leading to the activation of IFNA response genes GBP1, PIM1, CD40, IF130, IRF9, OAS1, IFIT2 and ISG15 show an increase in activity in PTB networks, and a consistent decrease in path activity over the course of treatment, as reflected by the reduction in node size and edge thickness. These sub-networks provide a representation of the interferon response in tuberculosis.

A detailed inspection of high activity paths in PTB_M reveals the presence of several cytokine mediated paths as well as several highly active paths leading to the expression of interferon-response genes. Many of the modulators of signalling such as JAK2, STAT1, and STAT3 are well known to be involved in regulation of the immune system. Figure 7 provides a list of few selected paths from a given source to a specified target that show considerably high activity in a given condition, and their variation in *Path Costs* in different conditions. Paths leading to antigen presentation through HLA-DRB5 are seen to be most repressed in PTB_M as compared to their activities in the HC_M top path network. Subversion of antigen presentation and expression of HLA genes is a survival strategy utilised by *M.tb*[52]. Toll like receptors play an important role in recognition and subsequent activation of immune responses to tuberculosis. Paths from TLR6 through MYD88 leading to TNF and IRF1 expression are active in disease, with the activity decreasing over treatment. Downstream responses of TNF through TNFRSF leading to GBP1 are also more pronounced in PTB, as opposed to the healthy controls. Analysis of top paths also reveals paths through the anti-inflammatory cytokine IL10 in networks for disease, reflective of the intricate interplay between both PICs and AICs. There are several other examples where paths revert to the HC_B paths upon treatment (Supplementary table S15). There are also examples of paths that do not completely revert their activities upon treatment, suggesting that treatment is effective in rescuing some but not all paths in the network (Supplementary table S16). From this analysis we thus identified the set of paths that respond to treatment and those that remain relatively unchanged over the course of treatment.

Discussion

In any cell, several proteins mediate a whole range of biochemical activities at any given time. Infection with *M.tb* triggers several changes to normal cellular activity. Some of these changes may be readily comprehensible with the available knowledge about the biology of *M.tb* infection, but the effect or reason of involvement of many other proteins are yet to be understood. Most often, only those proteins that are known to be important are focused upon leaving out a huge number of them unaddressed. Besides gaps in knowledge of this type, two major challenges in relating genotype to phenotype are a) the scale in terms of numbers of molecular players and b) the complexity in terms of a complex network of interactions that occur in a cell. A wholistic systems perspective is thus necessary to not only identify critical components but also to gain mechanistic insights into how they achieve a particular objective. Networks of protein-protein interactions capturing both structural (complex formation) and functional (genetic) associations provide a useful framework to address these aspects[53,54]. It is becoming increasingly feasible to construct large scale networks owing to increase in the availability of curated databases, improved cross-mapping of literature containing primary research data into annotation databases, parseable ontologies, and cross-accession gene identifiers. There have been many advances in network biology including improved and better accessible tools based on graph theoretical concepts to mine networks and to visualize them[55]. It is no surprise therefore that networks are increasingly being applied to study several complex issues including finding hubs, choke points, shortest paths and alternate routes[56]. Networks have their drawbacks as well, primarily because they present static snapshots of cellular activities at a given instant of time. To overcome this limitation, the concept of dynamic interactome or response networks have been introduced[18,57,58], where time or space-series data is superimposed onto the networks to gain insights about the changing patterns of communication within the network.

In this work, we report use of response networks in health and disease and also upon treatment to identify top most activities in the cell under different conditions. In a given condition, we may not know which paths in the network are active, let alone provide a rank list of most important activities. It is important therefore to mine all high activity paths in an unbiased fashion. Using the response networks, the problem of identifying high activities in a cell can be formulated as a problem of identifying shortest paths with least *Path Costs*. A path refers to a route through a connected set of nodes from any source to any sink. *Path Cost* refers to the cumulative score of individual edges in a path. Lower the cost of the path, higher is the

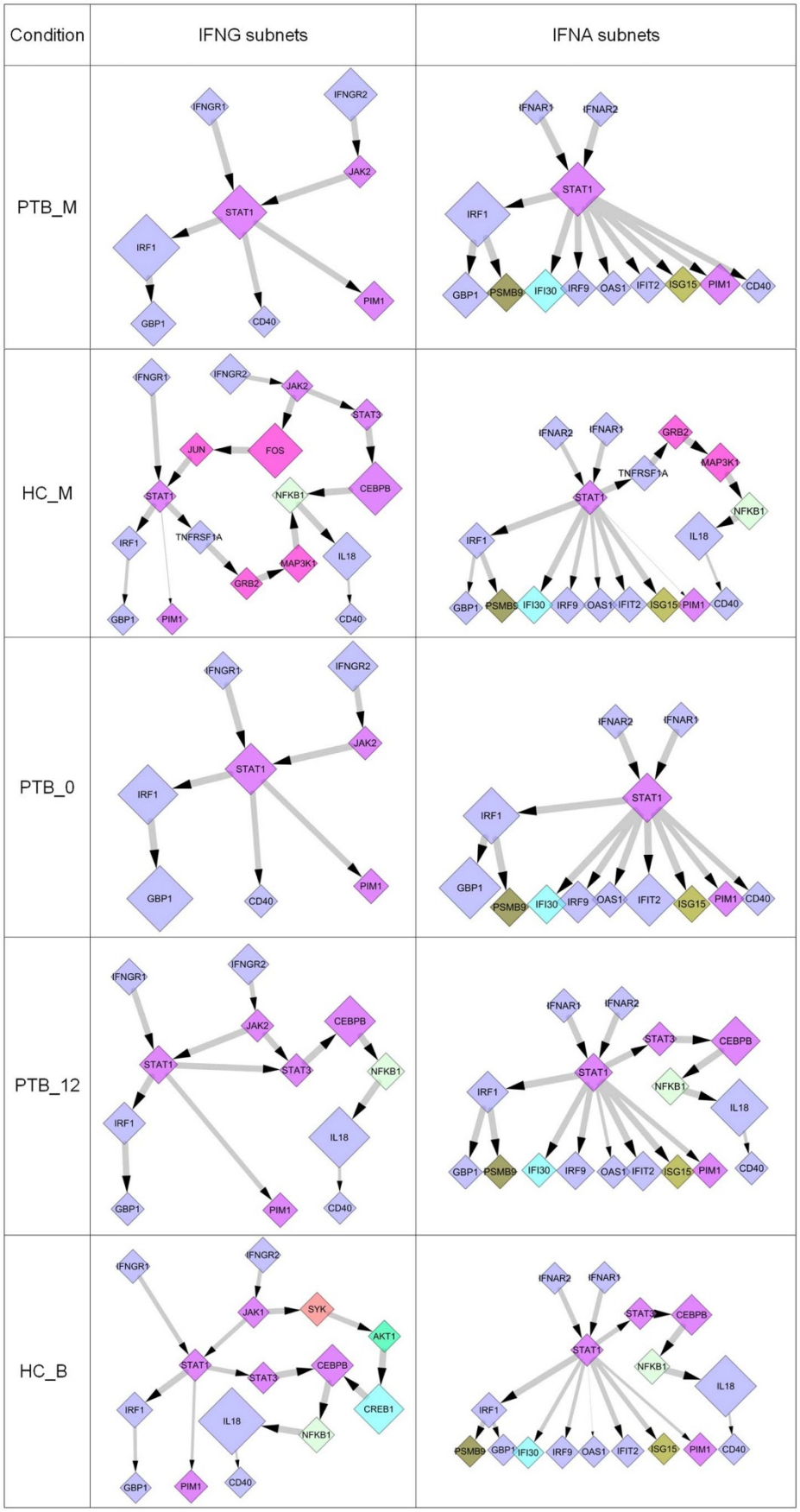


Figure 6 | Illustrative example of interferon sub-networks showing alterations in different conditions. Sub-networks are shown depicting paths from Interferon gamma receptors and Interferon alpha receptors to downstream target genes in different conditions. Node sizes correspond to node weights, while edge thickness reflects the interaction strength of the edge in a given condition. Nodes are coloured according to the modules they belong to.

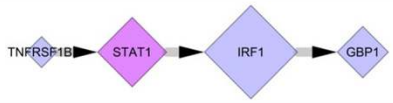
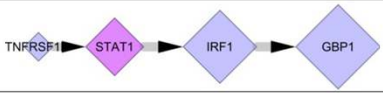
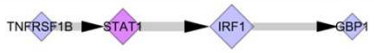
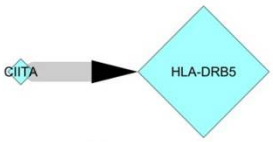
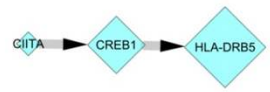
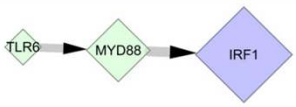
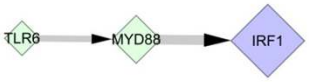
Source_target	Condition	Path Cost	Weighted Path
TNFRSF1B_GBP1	PTB_M	0.36	
	HC_M	5.04	
	PTB_0	0.42	
	PTB_12	1.46	
	HC_B	5.21	
CIITA_HLA-DRB5	PTB_M	237.68	
	HC_M	0.16	
	PTB_0	40.19	
	PTB_12	35.31	
	HC_B	2.03	
TLR6_TNF	PTB_M	1.06	
	HC_M	3.07	
	PTB_0	1.83	
	PTB_12	3.69	
	HC_B	4.51	
TLR6_IRF1	PTB_M	1.09	
	HC_M	3.16	
	PTB_0	1.29	
	PTB_12	2.71	
	HC_B	4.07	

Figure 7 | Path activities in different conditions. Example paths from a given source to target show variable *Path Costs* and thus variable activities in different conditions. The variation in activities are demonstrated by showing the differences in weights of nodes and edges constituting the path. Node sizes correspond to node weights, while edge thickness reflects the interaction strength of the edge in a given condition. Nodes are coloured according to the modules they belong to.

activity expected through it. This formulation enables identification of all possible paths of highest activities, anywhere in the cell. High activity paths are those with least *Path Costs* which implies sufficiently high levels of gene expression in that condition in all nodes of the path and also implicitly implies that the corresponding edges in the path are topologically centrally located. Hence the highest activity paths identified through this method can be expected to be significant routes through which intracellular communication can take place.

Often for studies of this nature, a limiting factor is utilisation of experimental data available in literature for the individual nodes in a comprehensive manner and reconstruction of a network. We have addressed that by first building a large curated network that captures known biology of the individual nodes in terms of their structural and function interactions with other proteins and includes direction of such interactions as well as the activatory or the inhibitory nature of the interactions. While care has been taken to include as much data as presently accessible, it is inevitable that the network remains incomplete due to lack of data for all nodes and all interactions. However, the network model presented here allows incorporation and refinement as new data becomes available.

Gene expression profiles through microarrays and recently also through NextGen sequencing provide lists of differentially regulated genes. Several reports presenting DEG profiles in tuberculosis patients exist in literature, leading to insights into the molecular map of the pathophysiology of tuberculosis. Many DEGs are consistently identified from individual studies leading to a general picture of what changes occur in response to tuberculosis. Examples of the identified profiles include (a) activation of cytokine signalling including both pro-inflammatory and anti-inflammatory cytokines (b) increased signalling through the Janus kinase signal transducer and activator of transcription (JAK-STAT) pathway (c) increased signalling through the Fc-gamma receptors and (d) Decrease in antigen presentation indicating subversion. The network analysis reported here identifies many of these DEGs to be present in the top most activity paths in the network, and more importantly also provides insights into the functional routes through which they can mediate their altered function. We note that over-expression of a node by itself is not a sufficient criterion for dictating 'High Activity' through a given path because for an activity to take place, the entire path must have a low Path Cost. Focusing on DEGs alone would miss out this aspect all together. Among our top activity paths, we find some nodes which are constitutively expressed in sufficient levels, but whose fold change may be close to 1, for example IL10 and IL18 are seen in high activity paths. IL10 and IL18 are both known to play critical roles in regulating the manifestation of tuberculosis[59,60]. We also observe some highly expressed DEGs such as AIM2, OSM, VAMP5 which do not have high scoring neighbours and hence do not feature in the top most paths. Placing them in a network perspective helps to understand their function in terms of their neighbourhoods. Analogous to finding highways in a complex traffic network, this approach finds top paths which are connected sets of nodes of sufficient weights (extent of expression) and high centrality.

Validation becomes an important step in any model building exercise. Since the network is built primarily using experimental data based on molecular level studies of individual proteins and weighted using experimentally derived gene expression profiles, validation will mainly amount to checking for consistency with well known biological phenomena. This is in addition to the model building validation steps that are routinely carried out involving data accession checks and cross-mapping across databases. Towards validation therefore, the following were analysed (a) known nodes appearing in the top paths, (b) checking that no inhibitory connections appear in the top paths (edges with inhibitory connections have nodes that are anti-correlated in their weights), (c) well known activated paths have sufficiently low *Path Costs* and known repressed paths have high costs and (d) path reversals upon treatment are seen in the treatment related response networks. All these criteria are met for the top paths that are analysed (Figures 6 and 7).

There are a large number of new insights that are gained from this study including identification of connected routes for well known nodes thus rationalising data from individual experiments. Thus, by integrating expression data and network topology, a comprehensive model of the host response to tuberculosis is presented. Mining the network identifies highest activities in the cell and interestingly identifies that such top paths all form a well connected subnet. Nodes in this network that differ substantially with respect to healthy control and in response to treatment have the potential to serve as possible biomarkers for tuberculosis and can be explored further for their diagnostic potential. The approach followed is sufficiently generic that it can be applied to other systems fairly easily, thus representing a general framework for integrating and rationalising genome-scale and clinical data.

Methods

The methodology comprises model building, network reconstruction, microarray data extraction and integration into the network to generate response networks, followed by a weighted shortest path analysis from all nodes to all other nodes in the network and computation of *Path Costs*.

Model reconstruction. In the first step towards generating a response network in macrophages, a model was built encompassing different processes that are activated in infection as well as those responsible for cellular signalling. The model was manually assembled after a thorough literature survey on the biology of the macrophage in tuberculosis. The model schematic was drawn using Microsoft Visio. Modules were identified through literature describing host cellular signalling events in tuberculosis, as well as from pathway information sources KEGG[61] and PATHWAYCOMMONS[62]. Once the signalling modules were identified, all proteins that participate in or belong to that module were enlisted using their ENTREZ identifiers.

Protein-protein interaction network. For all proteins identified in the model, corresponding interactions were extracted and a comprehensive protein-protein interaction network was constructed. Only interactions with experimental evidence and supporting literature were considered. The interactions were extracted from various sources including KEGG, STRING[63], NETPATH[64], Signalink[65] and from the macrophage interaction map curated by Raza and co-workers[66]. Through literature survey and network repositories such as BioGRAPH[67] and GeneMania[68], directions were assigned for several interactions by considering functional annotations such as 'phosphorylation', 'ubiquitinylation', 'proteolysis' and 'activation' as well as 'inhibition'. However, for certain interactions which describe a 'binding' event, the edge is represented as bidirectional. Interactions with experimental evidence but without functional annotation were also considered as bidirectional interactions. Ambiguous and predicted interactions were not considered.

Network Analysis. The network was subjected to network analysis using Cytoscape version 2.8.2[69]. Topological network properties were computed on the directed and unweighted network, using the NetworkAnalyzer[70] plugin. Shortest paths distribution was also computed between all vertices in the network. Edge centrality attributes, namely Edge Betweenness was calculated. Betweenness is a measure of centrality, which captures the number of times the edge is traversed among all shortest paths from the set of source nodes to the set of sink nodes[71]. For a graph G(V,E), the edge betweenness is defined as follows:

$$C_B(e) = \sum_{s \neq v \neq t \in E} \frac{s_{st}(e)}{s_{st}} \quad (1)$$

Where s_{st} is the number of shortest paths from s to t, and $s_{st}(e)$ is the number of shortest paths from s to t that pass through an edge e. The edge betweenness of each edge was computed using the Cytoscape plugin NetworkAnalyzer, where the betweenness is normalized by dividing by $(M-1)(M-2)$, where M is the number of edges in the connected component that the edge belongs to.

Microarray data integration. For this study, microarray expression data was taken from NCBI GEO ID GSE19491. The study was carried out by Berry et al. (2010)[8],where genome-wide whole blood transcriptional profiles were extracted from patients with active tuberculosis and monitored under drug treatment for 0, 2 and 12 months (5 samples each), as well as healthy controls (12 samples). Neutrophils, monocytes and CD4+ and CD8+ T cells were isolated. For this study, the monocyte expression data was considered for patients with active TB(7 samples) and healthy controls (4 samples).

Normalisation. The raw expression data was normalised per gene, as it was pre-normalised per-chip. The expression data for the genes of interest was filtered out. The average fold change of expression under conditions of infection in both whole

blood and isolated monocytes was computed with respect to the median value of the corresponding healthy controls. The fold change for a gene i in condition A with respect to its expression value in condition B was computed as described:

$$FC_{i(A/B)} = \frac{SignalIntensity_{i(A)}}{SignalIntensity_{i(B)}} \quad (2)$$

Where $SignalIntensity_{i(A)}$ represents the normalised signal intensity for the condition A and $SignalIntensity_{i(B)}$ represents the normalised signal intensity for the condition B.

Node weight. For all nodes in the network, the node weight was calculated as the product of the average fold change and the normalised signal intensity value for the corresponding node. This captures the quantitative information regarding expression level of the node in the cell.

The node weight $Ni(A)$ of node i in a given condition A, is computed as:

$$N_{i(A)} = FC_{i(A/B)} * SignalIntensity_{i(A)} \quad (3)$$

Edge weight. The edge weights for every edge between two nodes were calculated by taking into account both the node weights, and the betweenness of the given edge. The edge weight was computed as the inverse of the product of edge betweenness and the square root of the product of node weights. The edge weight in a given condition for an edge e ($W_{e\ (A)}$) comprised of nodes $N_{i(A)}$ and $N_{j(A)}$ is computed by the formula:

$$W_{e(A)} = \frac{1}{[C_B(e) * \sqrt{(N_{i(A)} * N_{j(A)})}]} \quad (4)$$

Where, $N_{i(A)}$ and $N_{j(A)}$ are the node weights of i and j nodes respectively in condition A. The edge weights were scaled up by 10^6, for convenience.

Module-module influence networks. Proteins were assigned into 32 modules based on their presence in a given KEGG Pathway. Modular node weights $N_{x(A)}$ for each module X were calculated as a cumulative sum of node weights of individual nodes (N_i) in the module.

$$N_{X(A)} = \sum N_{ix(A)} \quad (5)$$

Edge weight for an edge between module X and module Y in a condition was calculated as

$$W_{XY(A)} = \sum \frac{1}{W_{eXY(A)}} \quad (6)$$

where, $W_{eXY(A)}$ represents the edge weight W_e for an edge that captures an interaction between nodes belonging to module X and Y.

Shortest path computation. Shortest paths between all the proteins in the weighted and directed network were computed using Dijkstra's algorithm implemented in MATLAB-Boost Graph Library (http://www.stanford.edu/~dgleich/programs/matlab_bgl). The algorithm computes minimum weight shortest paths, in which each path begins from a source node and ends with a sink node, through interacting proteins, choosing the least-cost edge in every step. For a path of length n, the Path Cost was computed by taking the summation of the edge weights ($W_{e(A)}$) of all edges constituting the path.

$$PathCost = \sum_{e=1}^{n} W_{e(A)} \quad (7)$$

All paths were ranked according to their Path Cost, with the least costing paths ranked highest.

1. WHO Report No. WHO/HTM/TB/2012.6 (2012).
2. Casadevall, A. & Pirofski, L-a. Host-Pathogen Interactions: The Attributes of Virulence. *J. Infect Dis.* **184**, 337–344 (2001).
3. Mukherjee, S., Sambarey, A., Prashanthi, K. & Chandra, N. Current trends in modeling host–pathogen interactions. *Wiley Interdiscip Rev Data Min Knowl Discov.* **3**, 109–128 (2013).
4. Kaufmann, S. H. & Walker, B. Host–pathogen interactions. *Curr Opin Immunol.* **19**, 371–373 (2006).
5. Kirschner, D. E. & Linderman, J. J. Mathematical and computational approaches can complement experimental studies of host-pathogen interactions. *Cell Microbiol.* **11**, 531–539 (2009).
6. Raman, K., Bhat, A. G. & Chandra, N. A systems perspective of host-pathogen interactions. predicting disease outcome in tuberculosis. *Mol Biosyst.* **6**, 516–530 (2010).
7. Mandal, S., Sarkar, R. R. & Sinha, S. Mathematical models of malaria-a review. *Malar J.* **10**, 202 (2011).
8. Berry, M. P. R. *et al.* An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature.* **466**, 973–977 (2010).
9. Kumar, D. *et al.* Genome-wide Analysis of the Host Intracellular Network that Regulates Survival of Mycobacterium tuberculosis. *Cell.* **140**, 731–743 (2010).
10. Pommerenke, C. *et al.* Global Transcriptome Analysis in Influenza-Infected Mouse Lungs Reveals the Kinetics of Innate and Adaptive Host Immune Responses. *PLoS ONE.* **7**, e41169 (2012).
11. Núñez-Hernández, C. *et al.* Genome Expression Analysis of Nonproliferating Intracellular Salmonella enterica Serovar Typhimurium Unravels an Acid pH-Dependent PhoP-PhoQ Response Essential for Dormancy. *Infect Immun.* **81**, 154–165 (2012).
12. Nascimento, E. J. M. *et al.* Gene Expression Profiling during Early Acute Febrile Stage of Dengue Infection Can Predict the Disease Outcome. *PLoS ONE.* **4**, e7892 (2009).
13. Forst, C. V. Host-pathogen systems biology. *Drug Discov Today.* **11**, 220–227 (2006).
14. Barabasi, A-L. & Oltvai, Z. N. Network biology: understanding the cell's functional organization. *Nat Rev Genet.* **5**, 101–113 (2004).
15. Papin, J. A., Hunter, T., Palsson, B. O. & Subramaniam, S. Reconstruction of cellular signalling networks and analysis of their properties. *Nat Rev Mol Cell Biol.* **6**, 99–111 (2005).
16. Shoemaker, B. A. & Panchenko, A. R. Deciphering Protein-Protein Interactions. Part I. Experimental Techniques and Databases. *PLoS Comput Biol.* **3**, e42 (2007).
17. Shoemaker, B. A. & Panchenko, A. R. Deciphering Protein-Protein Interactions. Part II. Computational Methods to Predict Protein and Domain Interaction Partners. *PLoS Comput Biol.* **3**, e43 (2007).
18. Ideker, T., Ozier, O., Schwikowski, B. & Siegel, A. F. Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics.* **18**, S233–S240 (2002).
19. Steinman, R. M. Decisions about dendritic cells: past, present, and future. *Annu Rev Immunol* **30**, 1–22 (2012).
20. Urdahl, K. B., Shafiani, S. & Ernst, J. D. Initiation and regulation of T-cell responses in tuberculosis. *Mucosal Immunol* **4**, 288–93 (2011).
21. Andersen, P. Host responses and antigens involved in protective immunity to *Mycobacterium tuberculosis*. *Scand J Immunol.* **45**, 115–131 (1997).
22. Raja, A. Immunology of tuberculosis. *Indian J Med Res.* **120**, 213–232 (2004).
23. Kindt, T. J., Goldsby, R. A., Osborne, B. A. & Kuby, J. *Kuby Immunology: Sixth Edition* (2006).
24. Meraviglia, S. *et al.* γδ T Cells Cross-Link Innate and Adaptive Immunity in *Mycobacterium tuberculosis* Infection. *Clin Dev Immunol.* **2011** (2011).
25. O'Garra, A. *et al.* The Immune Response in Tuberculosis. *Annu Rev Immunol.* **31**, 475–527 (2013).
26. Ernst, J. D. Macrophage Receptors for Mycobacterium tuberculosis. *Infect Immun.* **66**, 1277–1281 (1998).
27. van Crevel, R., Ottenhoff, T. H. M. & van der Meer, J. W. M. Innate Immunity to *Mycobacterium tuberculosis*. *Clin Microbiol Rev.* **15**, 294–309 (2002).
28. Indik, Z., Kelly, C., Chien, P., Levinson, A. I. & Schreiberm, A. D. Human Fc gamma RII, in the absence of other Fc gamma receptors, mediates a phagocytic signal. *J. Clin. Invest.* **88**, 1766–1771 (1991).
29. Tuijnman, W. B., Capel, P. J. & van de Winkel, J. G. Human low-affinity IgG receptor Fc gamma RIIa (CD32) introduced into mouse fibroblasts mediates phagocytosis of sensitized erythrocytes. *Blood.* **79**, 1651–1656 (1992).
30. Stenmark, H., Botelhom, R. J., Scott, C. C. & Grinstein, S. Phosphoinositide Involvement in Phagocytosis and Phagosome Maturation. *Curr Top Microbiol Immunol.* **282**, 1–30 (2004).
31. Nunes,P. & Demaurex, N. The role of calcium signaling in phagocytosis. *J Leukoc Biol.* **88**, 57–68 (2010).
32. Rohde, K., Yates, R. M., Purdy, G. E. & Russell, D. G. *Mycobacterium tuberculosis* and the environment within the phagosome. *Immunol Rev.* **219**, 37–54 (2007).
33. Reiner, S. L. Helper T cell differentiation, inside and out. *Curr Opin Immunol.* **13**, 351–355 (2001).
34. Yamane, H. & Paul, W. E. Cytokines of the gamma(c) family control CD4+ T cell differentiation and function. *Nat Immunol.* **13**, 1037–1044 (2012).
35. Li, Z., Zhang, Y. & Sun, B. Current understanding of Th2 cell differentiation and function. *Protein Cell.* **2**, 604–611 (2011).
36. Cooper, A. M. & Khader, S. A. The role of cytokines in the initiation, expansion, and control of cellular immunity to tuberculosis. *Immunol Rev.* **226**, 191–204 (2008).
37. Kidd, P. Th1/Th2 balance: the hypothesis, its limitations, and implications for health and disease. *Altern Med Rev.* **8**, 223–246 (2003).
38. Schluger, N. W. & Rom, W. N. The host immune response to tuberculosis. *Am J Respir Crit Care Med.* **157**, 679–691 (1998).
39. Ding, Z., Xiong, K. & Issekutz, T. B. Chemokines stimulate human T lymphocyte transendothelial migration to utilize VLA-4 in addition to LFA-1. *J Leukoc Biol.* **69**, 458–466 (2001).
40. Astarie-Dequeker, C., Nigou, J., Puzo, G. & Maridonneau-Parini, I. Lipoarabinomannans activate the protein tyrosine kinase Hck in human neutrophils. *Infect Immun.* **68**, 4827–4830 (2000).
41. Tapping, R. I. & Tobias, P. S. Mycobacterial lipoarabinomannan mediates physical interactions between TLR1 and TLR2 to induce signaling. *J. Endotoxin Res.* **9**, 264–268 (2003).
42. Quesniaux, V. J. *et al.* Toll-Like Receptor 2 (TLR2)-Dependent-Positive and TLR2-Independent-Negative Regulation of Proinflammatory Cytokines by Mycobacterial Lipomannans. *J. Immunol.* **172**, 4425–4434 (2004).
43. Fratti, R. A., Chua, J., Vergnem,I. & Deretic, V. *Mycobacterium tuberculosis* glycosylated phosphatidylinositol causes phagosome maturation arrest. *PNAS.* **100**, 5437–5442 (2003).

44. Vergne, I. *et al. Mycobacterium tuberculosis* Phagosome Maturation Arrest: Mycobacterial Phosphatidylinositol Analog Phosphatidylinositol Mannoside Stimulates Early Endosomal Fusion. *Mol Biol Cell.* **15**, 751–760 (2004).
45. Spargo, B. J., Crowe, L. M., Ioneda, T., Beaman, B. L. & Crowe, J. H. Cord factor (alpha, alpha-trehalose 6,6′-dimycolate) inhibits fusion between phospholipid vesicles. *PNAS.* **88**, 737–740 (1991).
46. Rajaram, M. V. S. *et al. Mycobacterium tuberculosis* activates human macrophage peroxisome proliferator-activated receptor gamma linking mannose receptor recognition to regulation of immune responses. *J. Immunol.* **185**, 929–942 (2010).
47. Blumenthal, A. *et al.* The Wingless homolog WNT5A and its receptor Frizzled-5 regulate inflammatory responses of human mononuclear cells induced by microbial stimulation. *Blood.* **108**, 965–973 (2006).
48. Pereira, C., Schaer, D. J., Bachlim, E. B., Kurrer, M. O. & Schoedon, G. Wnt5A/CaMKII Signaling Contributes to the Inflammatory Response of Macrophages and Is a Target for the Antiinflammatory Action of Activated Protein C and Interleukin-10. *Arterioscler Thromb Vasc Biol.* **28**, 504–510 (2008).
49. Harding, C. V. & Boom, W. H. Regulation of antigen presentation by *Mycobacterium tuberculosis*: a role for Toll-like receptors. *Nat Rev Microbiol.* **8**, 296–307 (2010).
50. Zhu, X., Gerstein, M. & Snyder, M. Getting connected: analysis and principles of biological networks. *Genes Dev.* **21**, 1010–1024 (2007).
51. Bossers, K. *et al.* Intensity-based analysis of dual-color gene expression data as an alternative to ratio-based analysis to enhance reproducibility. *BMC Genomics.* **11**, 112 (2010).
52. Baena, A. & Porcelli, S. A. Evasion and subversion of antigen presentation by *Mycobacterium tuberculosis*. *Tissue Antigens.* **74**, 189–204 (2009).
53. De Las Rivas, J. & Fontanillo, C. Protein-protein interaction networks: unraveling the wiring of molecular machines within the cell. *Brief Funct Genomics.* **11**, 489–496 (2012).
54. Ma, X. & Gao, L. Biological network analysis: insights into structure and functions. *Brief Funct Genomics.* **11**, 434–442 (2012).
55. Lahti, L., Knuuttila, J. E. A. & Kaski, S. Global modeling of transcriptional responses in interaction networks. *Bioinformatics.* **26**, 2713–2720 (2010).
56. Jordán, F., Nguyen, T-P. & Liu, W-c. Studying protein-protein interaction networks: a systems view on diseases. *Brief Funct Genomics.* **11**, 497–504 (2012).
57. Ideker, T. & Sharan, R. Protein networks in disease. *Genome Res.* **18**, 644–652 (2008).
58. Davis, M. J., Shin, C. J., Jing, N. & Ragan, M. A. Rewiring the dynamic interactome. *Mol Biosyst.* **8**, 2054–2066 (2012).
59. Redford, P. S., Murray, P. J. & O'Garra, A. The role of IL-10 in immune regulation during M. tuberculosis infection. *Mucosal Immunol.* **4**, 261–270 (2011).
60. Schneider, B. E. *et al.* A role for IL-18 in protective immunity against *Mycobacterium tuberculosis*. *Eur J Immunol.* **40**, 396–405 (2010).
61. Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. & Tanabe, M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* **40**, D109–114 (2012).
62. Cerami, E. G. *et al.* Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res.* **39**, D685–690 (2010).
63. Szklarczyk, D. *et al.* The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.* **39**, D561–568 (2011).
64. Kandasamy, K. *et al.* NetPath: a public resource of curated signal transduction pathways. *Genome Biol.* **11**, R3 (2010).
65. Korcsmaros, T. *et al.* Uniformly curated signaling pathways reveal tissue-specific cross-talks and support drug target discovery. *Bioinformatics.* **26**, 2042–2050 (2010).
66. Raza,S. *et al.* Construction of a large scale integrated map of macrophage pathogen recognition and effector systems. *BMC Syst. Biol.* **4**, 63 (2010).
67. Liekens, A. *et al.* BioGraph: Unsupervised biomedical knowledge discovery via automated hypothesis generation. *Genome Biol.* **12**, R57 (2011).
68. Warde-Farley, D. *et al.* The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* **38**, W214–220 (2010).
69. Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P. L. & Ideker, T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics.* **27**, 431–432 (2010).
70. Assenov,Y., Ramirez, F., Schelhorn, S. E., Lengauer, T. & Albrecht, M. Computing topological parameters of biological networks. *Bioinformatics.* **24**, 282–284 (2008).
71. Linton, C. F. A Set of Measures of Centrality Based on Betweenness. *Sociometry.* **40**, 35–41 (1977).

Acknowledgements

We thank the Department of Biotechnology (DBT) and the Mathematical Biology Programme supported by the Department of Science and Technology (DST), Government of India, for financial support. The use of facilities at Department of Biochemistry, Indian Institute of Science is also gratefully acknowledged.

Author contributions

Both A.S. and K.P. contributed equally to the manuscript. A.S. and K.P. constructed the model and carried out the network analysis. N.S.C. conceptualised and closely supervised the project. All three authors read, wrote and approved of the manuscript.

Additional information

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Sambarey, A., Prashanthi, K. & Chandra, N. Mining large-scale response networks reveals 'topmost activities' in *Mycobacterium tuberculosis* infection. *Sci. Rep.* **3**, 2302; DOI:10.1038/srep02302 (2013).