

Differential co-expression network analysis for identification of functional modules in HIV infection and tuberculosis-HIV co-infection

Mohit Jha,¹ Anvita Gupta,¹ Sudha Singh,¹ Khushhali Menaria Pandey²

¹Department of Bioinformatics, Maulana Azad National Institute of Technology, Bhopal;

²Department of Biological Sciences & Engineering, Maulana Azad National Institute of Technology, Bhopal, India

Abstract

Co-infection with tuberculosis (TB) is the preeminent cause of demise in human immunodeficiency virus (HIV) infected individuals. However, diagnosis of TB, particularly in the presence of an HIV co-infection, can be limiting owing to the high inaccuracy associated with conventional diagnostic strategies. Here we determine dysregulated pathways in TB-HIV co-infection and HIV infection utilizing co-expression networks. Primarily, we utilized preservation statistics to identify gene modules that exhibit a weak conservation of network topology within HIV infected and TB-HIV co-infected networks. Raw data was downloaded from Gene Expression Omnibus (GSE50834) and duly pre-processed. Co-expression networks for each condition (HIV infected and TB-HIV co-infected) were constructed independently. Preservation of HIV infected network edges was evaluated with respect to TB-HIV co-infected and vice versa using weighted correlation network analysis. Two out of the 22 modules were identified as exhibiting weak preservation in both conditions. Functional enrichment analysis identified that weakly preserved modules were pertinent to the condition under study. For instance, weakly preserved TB-HIV co-infected module T1 enriched for genes associated with mito-

chondrion exhibited the highest fraction of gene interaction pairs exclusive to TB-HIV co-infection. Concisely, we illustrated the application of using preservation statistics to detect modules functionally linked with dysregulated pathways in disease, as exemplified by the mitochondrion module T1. Our analyses discovered gene clusters that are non-randomly linked with the disease. Highly specific gene pairs pointed to interactions between known markers of disease and favoured identification of possible markers that are likely to be associated with the disease.

Introduction

Human immunodeficiency virus (HIV) allied tuberculosis (TB) continues to be a leading global public health issue. Co-infection with TB and HIV has been closely linked since the emergence of AIDS. HIV infection has put up a significant rise in the global incidence of TB. In tandem, these infections – 1 viral and 1 bacterial, 1 newly emerged and 1 bygone – kill almost 4 million individuals every year, most of whom live in developing nations.¹

Co-infection with *Mycobacterium Tuberculosis (Mtb)* and HIV (TB-HIV) leads to challenges in diagnosis as well as treatment of tuberculosis. The mere presence of HIV makes an individual more vulnerable to developing TB disease, and having TB disease speed up HIV disease progression. The debilitation of the host's immune system by HIV increases possibility of *de-novo* co-infection with TB, or reactivation of latent TB.²⁻⁴ It causes the two pathogens to interact synergistically, assisting the progression of disease by a progressive decline in cell-mediated immunity. HIV alters the pathogenesis of TB, substantially raising the chances of disease from TB in HIV co-infected individuals and leading to continual extra pulmonary involvement, atypical radiographic expression, and paucibacillary disease, which can hinder timely diagnosis.

Here we present a different perspective to tuberculosis research by utilizing a co-expression network driven approach to gather insights into molecular interactions dysregulated in TB-HIV co-infected with respect to HIV infected and vice versa. Co-expression networks are being progressively used for understanding disease mechanisms and providing system - level outlook of dysregulated pathways.^{5,6} The basic presumption of co-expression analysis is that genes which exhibit strong correlation are more likely to be functionally related. Weighted Gene Correlation Network Analysis (WGCNA) performs co-expression analysis applying a network theory approach. It integrates expression differences across samples into a higher order network structure, revealing biologically meaningful relationships among genes based on their co-expression profiles.^{7,8} Here, we utilize a set of preservation statistics, to explore global differences in disease mechanisms under-

Correspondence: Khushhali Menaria Pandey, Department of Biological Sciences & Engineering, Maulana Azad National Institute of Technology, Bhopal, India.

E-mail: khushhalimenaria@manit.ac.in

Keywords: HIV; Tuberculosis; Co-expression network; Preservation statistics; Differential network analysis.

Contributions: the authors contributed equally.

Conflict of interest: the authors declare no potential conflict of interest.

Received for publication: 21 April 2016.

Revision received: 28 August 2016.

Accepted for publication: 15 September 2016.

©Copyright M. Jha et al., 2016

Licensee PAGEPress, Italy

Journal of Biological Research 2016; 89:5950

doi:10.4081/jbr.2016.5950

This article is distributed under the terms of the Creative Commons Attribution Noncommercial License (by-nc 4.0) which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

lying TB-HIV and HIV.⁹ Conventionally, these statistics have been utilized to identify gene modules that are topologically preserved between networks. Contrary to this approach, here we utilize these preservation statistics to identify gene modules that do not show topological preservation between networks. This is based on the presumption that such modules would represent a set of gene interactions that are specific to a condition and point to dysfunctional pathways and interactions.

In this study, we analyzed previously published gene expression data from individuals that were either mono-infected with HIV or co-infected with HIV and TB.¹⁰ We evaluated differential mechanisms between TB-HIV and HIV infected constructing co-expression networks independently for TB-HIV and HIV then clustering each of the co-expression networks to form groups of biologically relevant gene modules for each condition and finally preservation of modular topology from one condition was discovered with respect to the other condition. This approach allowed us to identify differences in gene connectivity patterns between conditions. The outcome of our differential co-expression analysis reveals convergent molecular mechanisms that are consistent with previously published studies and present novel hypothesis on gene interactions associated with these conditions.

Materials and Methods

Data acquisition

The dataset used in the current work were downloaded from the National Center for Biotechnology Information Gene Expression Omnibus data repository (<http://www.ncbi.nlm.nih.gov/geo/>) as gene expression series matrix files and were read into R statistical package using the GEOquery package.¹⁰⁻¹³ The primary results were based on dataset GSE50834.¹⁰ This data consists of 44 Illumina HumanHT-12 V4.0 expression beadchip with 23 samples HIV infected, 21 samples of TB-HIV co-infected. Two datasets GSE19439 and GSE19444 were used to eliminate false positive interactions.¹⁴ A summary of these two datasets is described in the Appendix.

Data pre-processing

All probes with missing Entrez gene identifiers were excluded from this study and multiple probes were collapsed into a single unique gene based on variance, resulting in a final reduced expression data set comprising of 11929 genes and 44 samples. The entire dataset is divided into two datasets containing HIV infected and TB-HIV co-infected samples, respectively. The two datasets were individually quantile-normalized and log transformed. The datasets thus obtained were sufficiently small for co-expression networks construction.

Weighted gene correlation network analysis

WGCNA is a systems-biology analysis method to identify and characterize gene modules whose members share strong co-expression. This method relies on constructing a weighted adjacency using Pearson correlations:

$$a_{ij} = |c_{ij}|^\beta \quad (\text{eq. 1})$$

where c_{ij} is Pearson correlation between gene i and gene j and a_{ij} is adjacency between gene i and gene j . β is a soft-thresholding parameter that gives prominence on stronger correlations over weaker and less meaningful ones while preserving the continuous nature of gene-gene relationships. β was selected in this analysis based on the criterion detailed by Zhang *et al.*⁸ The topological overlap-based matrix was then computed from the weighted adjacency matrix.^{15,16}

$$TOM_{ij} = \frac{\sum_u (a_{iu}a_{uj}) + a_{ij}}{\min\{k_i, k_j\} + 1 - a_{ij}} \quad (\text{eq. 2})$$

where index u runs across all nodes of the network, $k_i = \sum_u a_{iu}$ represents the total connectivity of gene i with all other genes in the network. The topological overlap reflects the relative interconnectedness (*i.e.* shared neighbors) between two genes.⁸

Topological overlap measure takes values in between 0 and 1, where 0 indicates no similarity between the gene pair while 1 indicates a direct link. Adjacency matrices of TB-HIV co-infected and HIV infected samples were taken as input to *TOMsimilarity* function available via WGCNA to generate co-expression networks. Hierarchical clustering on the topological dissimilarity (1-TOM) was executed using the function *flashClust*. Appendix provides a list of all genes identified in each of the networks (TB-HIV co-infected and HIV infected respectively) and their corresponding module assignments.

The construction of the network dendrogram based on this dissimilarity measure enables for the identification of gene modules whose associate members share strong interconnectivity patterns. The WGCNA *cutreeHybrid* R function was utilized to single out a appropriate cut height for module identification via an adaptive cut height selection approach.^{17,18} Gene modules, represented as branches of the network dendrogram, were assigned colors for visualization.

Preservation of modules

Module preservation statistics allows us to evaluate the conservation of gene pairs amid two networks (test and reference).⁹ Briefly, three types of network based module preservation statistics have been detailed by this method, namely: i) density based preservation statistics: four independent measures account for this statistic (these statistics determine if genes remain highly connected in the test network); ii) connectivity based preservation statistics: three independent measures of the network account for this statistic (they determine the extent to which connectivity patterns between genes in the reference network are similar to the genes in test network); iii) separability based preservation statistics: these determine if network modules remain distinct from one another in the test network.

In contrast to several existing methods that ascertain module preservation, network based statistics utilized by WGCNA does not require identification of gene modules within the test network to discover the conservation of reference modules within the test network. The authors of the original publication have successfully established that this method can identify sets of preserved co-expression modules.

To measure distinct aspects of module preservation, combination of two or more individual measures in a single measure that result in a single score is defined.

Median rank is a composite measure that is based on observed preservation values and does not need permutation testing. medianRank is the mean of median ranks computed for connectivity and density measures of each network module $[(\text{medianRank}_{\text{Connectivity}} + \text{medianRank}_{\text{Density}})/2]$. Usually, it shows no dependence on the size of network module.

Z_{Summary} is a permutation-based composite Z statistic measure that is used to evaluate the importance of observed statistics and is defined as the mean of Z scores computed for density and connectivity measures $[(Z_{\text{Density}} + Z_{\text{Connectivity}})/2]$.

The median rank is utilized to identify module preservation whereas Z_{Summary} is employed to assess significance of module preservation by permutation testing. A median rank of 10 was chosen as a cutoff to detect weak preservation with respect to the number of modules within each of the networks. Permutation was performed 200 times in accordance with the computational complexity required for our network

sizes. Based on the thresholds prescribed,⁹ modules with a Z_{Summary} score >10 indicate strong preservation, 2 to 10 specify weak to moderate preservation and <2 point to no preservation.

Condition specific gene pairs

Condition specific interactions for a given pair of genes i and j was defined as:¹⁹

$$\text{specificity}_{\text{condition } 1} = \frac{TOM_{ij(\text{condition } 1)}}{TOM_{ij(\text{condition } 1)} + TOM_{ij(\text{condition } 2)}} \quad (\text{eq. 3})$$

where, $TOM_{ij(\text{condition})}$ is the normalized topological overlap for the gene pairs i - j in the given condition (HIV infected or TB-HIV co-infected). A gene pair is considered to be condition specific, if the specificity was >0.80 and was in the top 1% of the gene pairs ranked on TOM similarity in a particular module. Considering the top 1% gene pairs allowed us to emphasize only on the strongest co-expression patterns within the module.

Functional enrichment analysis and visualization

The results presented correspond to the top term in the highest-ranking cluster using the annotation clustering feature of DAVID,²⁰ with Gene Ontology's functional annotations. Appendix provides functional annotation clusters identified for each of the modules within the TB-HIV co-infected and HIV infected networks respectively.

Availability of supporting data

The datasets supporting the results of this research article are available through the Gene expression omnibus database repository^{10,13,14}

Results and Discussion

Co-expression network construction and modularity detection

WGCNA was utilized to construct signed weighted co-expression networks from 21 HIV infected and 22 TB-HIV samples across 11929 genes. Briefly, correlation matrices were raised to a power $\beta = 9$ to obtain signed network adjacency matrices.⁸ The network adjacencies were

transformed to similarity matrices for subsequent clustering. In a network, each entry of the similarity matrix ij denotes the weight on the edge between genes. The degree of similarity between genes depends on the shared network neighborhood and their correlation.⁸ Clustering based on such a similarity measure allowed for identification of gene sets that were biologically meaningful.

Hierarchical clustering of the two weighted networks resulted in 22 modules for the network from HIV infected samples and 22 modules from the network derived from TB-HIV co-infected samples. Appendix presents the gene dendrograms and corresponding modules in both the networks. Genes that did not cluster were eliminated from further analysis.

Functional classification of modules identified in HIV infected and tuberculosis and HIV co-infected networks

Modules discovered by WGCNA have been shown to be biologically relevant to the condition under study.^{6,21} We utilized functional enrichment analysis to assess the functional significance of modules identified within each of the networks.

Classifying modules of the tuberculosis and HIV co-infected network

Enrichment of modules from the TB-HIV co-infected network revealed functions that are particularly associated with TB-HIV co-infection like mitochondrion, endosome transport, translational elongation, zinc ion binding, regulation of transcription, cytoskeletal part, antioxidant activity, plasma membrane part, mRNA metabolic process, GTPase regulator activity (Table 1).

Mtb infection either induces or inhibits host cell death, depending on the cell microenvironment and the bacterial strain. There is evidence indicating a role for mitochondria in these processes. *Mycobacterium* causes mitochondrial outer membrane permeabilization characterized by cytochrome *c* release from the mitochondrial inter-membrane space and apoptosis. Duan *et al.* (2002) and Chen *et al.* (2006) have reported that mycobacterial infection affects mitochondrial function.^{22,23} There is consistent evidence for a central role of mitochondrial dysfunction in these pathologies.²⁴ Nucleotide reverse transcriptase inhibitors (NRTIs) have been described to be primarily responsible for mitochondrial dysfunction in adipose tissue and liver. Several mechanisms explain these deleterious effects of highly active

Table 1. Enrichment of modules identified in the tuberculosis and HIV co-infected network and HIV network.

TB-HIV network				HIV network			
Module	#Nodes	Top term	P	Module	#Nodes	Top term	P
T1	465	Mitochondrion	4.71E-16	H1	223	Modification-dependent protein catabolic process	2.96E-03
T2	307	Endosome transport	2.53E-03	H2	235	Nucleoplasm	3.07E-05
T3	728	Translational elongation	1.33E-57	H3	575	Eukaryotic translation initiation factor 4F complex	1.27E-03
T4	290	Zinc ion binding	0.0049276	H4	252	Exocytosis	1.35E-02
T5	501	Regulation of transcription	4.654E-06	H5	401	Transcription	1.00E-07
T6	356	Cytoskeletal part	0.0001582	H6	134	Cell cycle phase	1.43E-05
T7	179	Antioxidant activity	0.0014922	H7	203	Regulation of locomotion	2.93E-03
T8	216	Plasma membrane part	0.0022456	H8	224	Nuclear lumen	0.016861
T9	426	mRNA metabolic process	0.0006607	H9	686	Microbody	0.0028034
T10	444	GTPase regulator activity	0.0021437	H10	262	Response to wounding	0.0008352
T11	258	Spermatid development	0.0026576	H11	158	tRNA processing	0.0095064

TB-HIV, tuberculosis and human immunodeficiency virus co-infected; HIV, human immunodeficiency virus infected.

antiretroviral therapy on mitochondria, and evidence points to other mechanisms beyond the *Pol-hypothesis*. HIV infection also has direct effects on mitochondria. According to Cossarizza *et al.* (1998) the increase in tumor necrosis factor (TNF) production,²⁵ is known to directly affect the mitochondrial function in patients with HIV infection. For instance, T1 module contained several genes associated with mitochondria and respiratory electron transport chain such as *COX5A*, *COX5B*, *NDUFB8*, *NDUFB6*, *NDUFB2*. T2 module was enriched with genes known to participate in endosomal activity. It has been concluded that Alveolar macrophages from patients with tuberculosis and HIV infections were competent to endocytose and phagocytose material and to deliver the material to functional, acidified lysosomes. *Mtb* resides in non-acidified vacuoles in endocytically Competent Alveolar Macrophages from patients with Tuberculosis and HIV Infection co-infection.²⁶

Classifying modules of the HIV infected network

Though the same set of 11929 genes were used to construct the network in each condition, modules cluster distinctly based on their co-expression. Enrichment of modules from the HIV infected network revealed several biological functions typically associated with HIV infection such as protein catabolic process, exocytosis, transcription, regulation of locomotion, nuclear lumen, response to wounding, tRNA processing, transition metal ion binding, translational elongation (Table 1).

Past studies have suggested that HIV infection is associated with increased protein catabolism and that protease inhibitor therapy improves lean body mass (LBM). Several genes associated with the protein catabolic activity were found in module H1 such as *UBXN1*, *SPSB1*, *HECTD3*, *FBXW2*, *PSMD5*, and *FBXO21*.

The mediator complex is regarded as a global regulator of gene expression. It may function in distinct ways with different genes or different cell types, including those susceptible to HIV infection,^{27,28} with a high degree of structural flexibility, and variable subunit composition, representing a general requirement (*e.g.* enhancer-driven) for activated RNAP II transcription.^{29,30} The role of some Mediator complex subunits in HIV replication has been attributed to a role on viral transcription, due to its RNAP II-related function or its link to the Tat-mediated transcription through the HIV long terminal repeat.^{27,31} For instance, H2 contained several genes associated with mediator complex such as *MED13L*, *MED13*, *MED9*, and *MED30*. Appendix provides results of enrichment analysis in each of the networks (TBHIV co-infected and HIV infected respectively) and their respective modules.

A systems approach to identify functional distinctness between tuberculosis and HIV and HIV

The functional annotation clustering of the modules reveals an exclusivity of certain functions between TB-HIV co-infected and HIV infected individuals suggesting distinctness in the network topology of connections for genes within these networks. In order to evaluate and quantify differential gene co-expression, we performed a network preservation analysis. This approach enabled us to identify modules that are adequately unique in terms of their gene co-expression within a given network, in comparison to other network. We utilized a function implemented in WGCNA called modulePreservation.⁹ In contrast to the viewpoint of the original paper that intends to identify highly preserved modules between disease conditions, we intend to identify weakly preserved modules across conditions. We hypothesized that network modules that were weakly preserved in either condition may reveal dysregulated pathways or interactions in disease that were either gained or lost with respect to a particular condition under study.

Assessing differential co-expression in tuberculosis and HIV with respect to HIV

In order to evaluate how the topology of the TB-HIV network differed from the HIV network, we calculated the preservation (density, connectivity and separability) statistics of modules from the TB-HIV infected network (reference network) as compared to the HIV co-infected network (test network). Lower preservation of modules points to a loss of co-expression structure between these gene pairs in the TB-HIV co-infected network. Appendix presents a table of observed preservation statistics for all modules of the TB-HIV co-infected network. The Z_{Summary} scores likewise revealed a low preservation of these modules via permutation testing (Table 2).

TB-HIV co-infected, in general, exhibits profound effect on the immune system, with pathways associated with wounding and inflammation and defense response being subsequently activated. The functional enrichment identified within modules (Table 1) validated our approach, underlining functions that are more pronounced in TB-HIV compared to HIV such as mitochondria, electron transport chain and endosome. These modules also exhibited higher specificity of connections to the TB-HIV network than their preserved counterparts (Table 3).

For instance, the highest specificity was observed for module T1 with nearly 23% of its gene pairs as being specific to TB-HIV (specificity >

Table 2. Permutation-based Z_{summary} in tuberculosis and HIV network and HIV network (medianRank>10).

Module	TB-HIV network			Module	HIV network		
	Size	Z_{Summary}	MedianRank		Size	Z_{Summary}	MedianRank
T1	465	4.52	22	H1	223	7.79	22
T2	307	8.86	20	H2	235	9.34	21
T3	728	16.22	20	H3	575	16.79	20
T4	290	10.79	19	H4	252	14.44	20
T5	501	18.98	18	H5	401	16.69	18
T6	356	17.11	17	H6	134	10.16	16
T7	179	11.13	16	H7	203	14.47	16
T8	216	14.63	15	H8	224	15.40	15
T9	426	18.83	15	H9	686	28.43	13
T10	444	20.79	13	H10	262	18.95	12
T11	258	18.85	11	H11	158	14.96	12

TB-HIV, tuberculosis and human immunodeficiency virus co-infected; HIV, human immunodeficiency virus infected.

0.80). A table of highly specific gene pairs for all modules of the TB-HIV co-infected network is provided in the Appendix.

Interestingly, the TB-HIV-specific gene interactions identified in module T1 corresponded with interactions categorized as a part of the mitochondrion and respiratory transport chain (Figure 1a). For instance, expression of *SRP14P1*, an Alu RNA binding protein (also called *Signal Recognition Particle 14 kDa Pseudogene1*), regulates the expression of Alu elements, which is an important expression regulation element. It has potential role in affecting the pathogenesis of HIV infection.³² Additionally, *SRP14P1* shows high specificity interactions with certain mitochondrial genes including *NDUFA3*, *COX5B*, *NDUFS5* (Figure 1a). *PIPSL*, which binds to ubiquitinated proteins, is strongly co-expressed within module T1. It shares highly specific and highly correlated gene neighbors with mitochondrial gene *LOC402175*, such as *IDH3B*, *TMEM141*, *PSME3*, *POP7*, *SLBP*, *HLA-B*, *ETFB*, and *COMMD5*. Though the exact mechanism in TB-HIV co-infection is not well understood, there is evidence to suggest that they are significantly related (Figure 1a).

VAMP7, a protein-coding gene (also called *vesicle associated membrane protein 7*), encodes a transmembrane protein that is a member of the soluble N-ethylmaleimide-sensitive factor attachment protein receptor (SNARE) family. The encoded protein localizes to late endosomes and lysosomes and is involved in the fusion of transport vesicles to their target membranes. Similarly *VPS54*, a protein coding gene (also called *Vacuolar protein sorting-associated protein 54*), encodes for a protein that in yeast forms part of a trimeric vacuolar-protein-sorting complex which is required for retrograde transport of proteins from prevacuoles to the late Golgi compartment. Both *VAMP7* and *VPS54* are highly specific and strongly co-expressed within module T2.

Though there is no direct evidence for the interaction of these markers, the exact understanding of these markers combined with the high specificity of co-expression observed in our network module suggests possible associations TB-HIV co-infection. Overall, our results indicate that the modules exhibiting low preservation contain a number of gene pairs that are likely to be linked with the disease progression. Though it is likely that not all the genes within these less-preserved modules play a significant role in disease, a number of highly specific gene pairs identified were noted to play a role in TB-HIV.

Assessing differential co-expression in HIV with respect to tuberculosis and HIV

A similar analysis with HIV network as the reference network

enabled us to identify gene pairs that were not preserved in TB-HIV network. As proposed earlier, we speculated that identifying non - / weakly - preserved modules in the HIV network could point to gene associations that are *gained* in TB-HIV (Figure 1c). We identified two modules- H1 and H2 that exhibited weak preservation in the TB-HIV network.

Based on the median preservation score and Z_{summary} , we identified 2 interesting modules. Modules H1 and H2 from the HIV network were weakly preserved in the TB-HIV network (Figure 1d). A table of the observed preservation statistics for all modules of the HIV infected network is provided in the Appendix. Z_{summary} , a permutation statistic that assess significance of the observed preservation, also exhibit a low preservation of these network modules (Table 2).

FBXO21 (F-Box Protein 21), a protein coding gene, is strongly co-expressed within module H1 and shares highly specific and highly correlated gene neighbors with *UBXN1*, such as *SPSB1*, *PTPN7*, *RBM23*, *KIAA1967*, *DIABLO*, *PRPF3*. Interestingly, both *UBXN1* and *FBXO21* are involved in protein catabolic process (Figure 1c). Similarly *SMNDC1*, a protein-coding gene (also called *Survival Motor Neuron Domain Containing 1*), is strongly co-expressed in module H2. GO annotations reveals its role in regulation of apoptosis.

We utilized a co-expression specificity measure to elucidate co-expressed genes pairs from these 2 modules.¹⁹ We observed that modules showing low preservation in the HIV infected network consistently had a higher proportion of gene pairs exclusive to the HIV infected network than their preserved counterparts (Table 3). For instance, ~27% of the gene pairs considered within H1 and ~20% of the gene pairs considered within H2 were specific to the HIV infected network while. The exclusivity of connections to the HIV infected network further emphasizes the possibility of a role in HIV disease progression. A table of highly specific gene interaction pairs for all modules of the HIV infected network is provided in the Appendix.

Identification of false positive gene interactions

To identify the false positive interactions we constructed co-expression networks for latent TB and pulmonary TB independently. Clustering each network resulted in sets of correlated genes (modules) for each condition. We utilized network preservation statistics to identify modules that do not exhibit a preservation of topology with healthy network. Such modules would represent a set of gene interactions that

Table 3. Fraction of edges exclusive to the tuberculosis and HIV as compared to the HIV network and vice versa.

Module	Size	TB-HIV network			Module	Size	HIV network		
		Gene pair above cutoff	Top 1% gene pairs	Specificity (%)			Gene pairs above cutoff	Top 1% gene pairs	Specificity (%)
T1	465	249	1079	23.08	H1	223	67	248	27.02
T2	307	118	470	25.11	H2	235	57	275	20.73
T3	728	313	2646	11.83	H3	575	150	1650	9.09
T4	290	53	419	12.65	H4	252	52	316	16.46
T5	501	148	1253	11.81	H5	401	114	802	14.21
T6	356	84	632	13.29	H6	134	22	89	24.72
T7	179	11	159	6.92	H7	203	6	205	2.93
T8	216	2	232	0.86	H8	224	25	250	10.00
T9	426	56	905	6.19	H9	686	109	2350	4.64
T10	444	62	983	6.31	H10	262	16	342	4.68
T11	258	15	332	4.52	H11	158	4	124	3.23

TB-HIV, tuberculosis and human immunodeficiency virus co-infected; HIV, human immunodeficiency virus infected.

exists in LTB/PTB but not healthy. Details of the procedure used is provided in the Appendix.

On the basis of median rank preservation and Z_{Summary} , we identified two latent TB modules LTB5 and LTB6 and three pulmonary TB modules PTB7, PTB8 and PTB9 that exhibited weak preservation in the healthy network. A table showing module preservation in latent and pulmonary TB is provided in the Appendix. The results of functional annotation clustering indicated that certain functions are mutual exclusive between pulmonary TB and latent TB implying a difference in the topology of connections for genes within these networks. The Appendix provides results of enrichment analysis in each of the networks and their respective modules.

Latent TB specific gene interactions identified in unpreserved module LTB5 corresponded with interactions classified as a part of the phosphorylation process (*ATP6V1C1*, *KISS1R*, *CLK4*, *PRKDC*), cytoskeleton (*NDE1*, *MAP7*, *C16ORF48*) & regulation of transcription (*ZBTB34*, *GMEB2*, *ZXDC*, *PRKDC*, *LRCHA*) while unpreserved module

LTB6 corresponded with interactions classified as a part of inflammatory response, defense response, and response to wounding (*CD97*, *CD46*, *CSF3R*, *STAT3*).

Pulmonary TB specific gene interactions identified in unpreserved module PTB7 where *FAIM3* is strongly co-expressed corresponded with interactions classified as a part of plasma membrane (*FAIM3*, *MAL*, *SLC38A1*, *ITPR3*), and regulation of apoptosis (*FAIM3*, *MALT1*, *MAL*), whereas in PTB8 where *ANKRD22*, *FBXO6* and *GBP1* are strongly co-expressed, interactions are classified as a part of intracellular signaling cascade (*IRAK2*, *PLEK*, *GADD45G*, *FBXO6*) and protein kinase cascade (*IRAK2*, *PLEK*, *GADD45G*). Similarly, pulmonary TB specific gene interactions identified in unpreserved module PTB9, where *C16ORF57* is strongly co-expressed, corresponded with gene interactions classified as a part of immune response (*TNFAIP8L2*, *IFITM2*, *NFIL3*, *CHUK*), transmembrane region (*IFITM2*, *MBOAT1*, *IL17RA*) and polymorphism (*SMCHD1*, *PDLIM7*, *SMARCD3*, *IFITM2*, *MBOAT1*, *CTSD*, *C16ORF57*, *CHUK*, *IL17RA*, *OS9*). A table of the module specific gene interaction

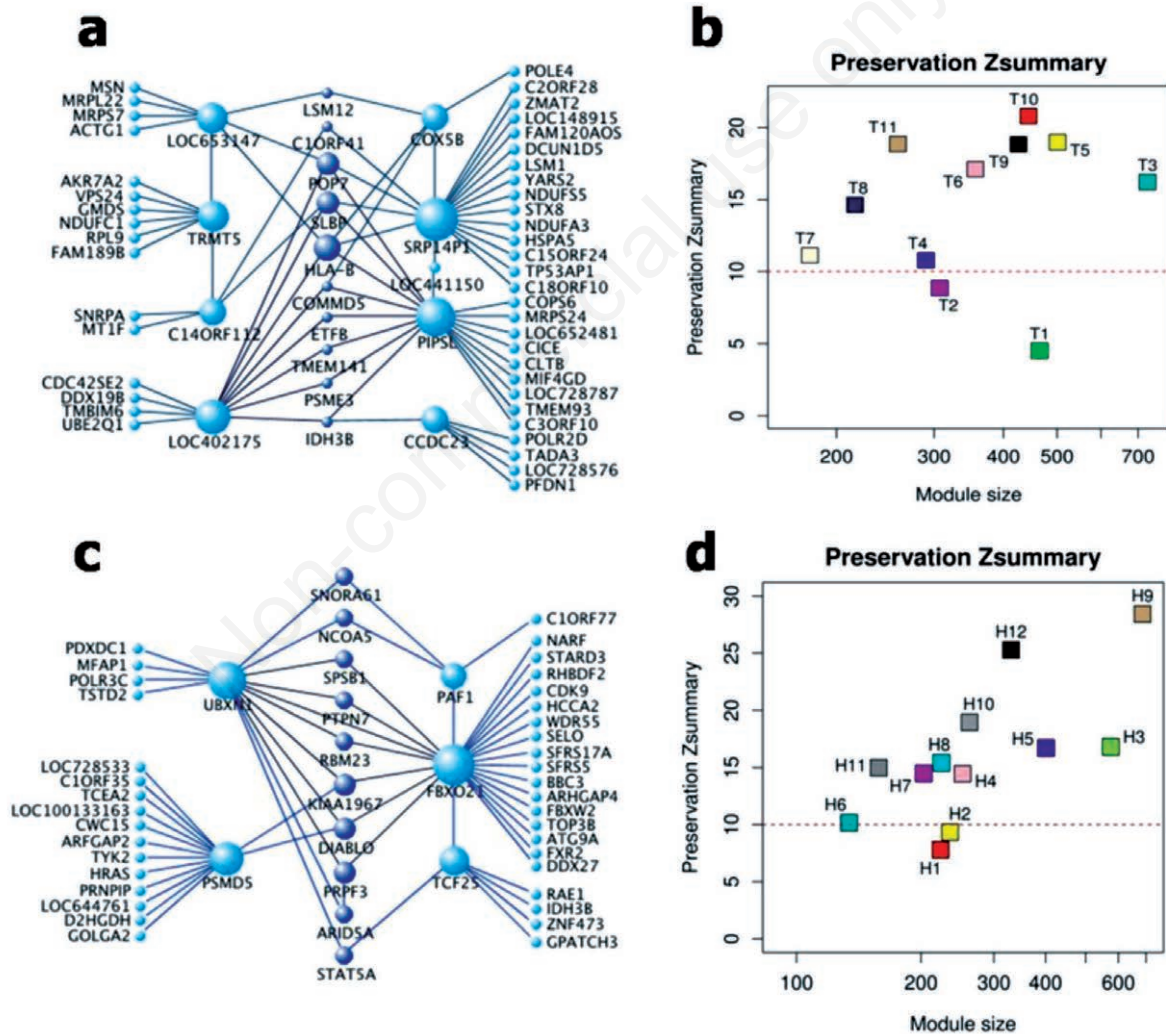


Figure 1. a) Differential co-expression in tuberculosis and human immunodeficiency virus co-infected (TB-HIV) with respect to HIV. Co-expression pattern between genes identified in module T1 are shown here. All interactions have a specificity and correlation >0.80 . The size of the node is proportional to the number of edges incident on that node. b) Scatter plot identifying the Z_{Summary} of module preservation between reference (TB-HIV) and test (HIV) networks. c) Differential co-expression in HIV network with respect to TB-HIV network, co-expression pattern between a subset of genes identified in module H1 from the HIV network. The size of the node is proportional to the degree of the node in the network shown. d) Scatter plot identifying the Z_{Summary} of module preservation between reference (HIV) networks and test (TB-HIV) network.

pairs for all modules of the latent TB and pulmonary TB network is provided in the Appendix. Comparative analysis of the condition specific gene interactions found in the respective modules reveals that there are no false positive gene interactions in our study.

Conclusions

The network preservation analysis indicates lack of similarity for some of the modules. A systematic evaluation of gene modules having a low preservation between TB-HIV co-infected and HIV infected conditions showed much higher specificity among gene pairs relevant to the condition under study. Based on functional analysis of data, we identified distinct sets of co-expressed genes related to biological gene sets. In addition, we distinguished between preserved and non-preserved modules between TB-HIV co-infected and HIV infected. We illustrated the application of using network preservation statistics to detect modules functionally associated with dysregulated pathways in disease, as exemplified by the mitochondrion module T1. In summary, this approach enabled identifying putative biomarkers that are likely to be associated with the disease and can be used to identify differences between disease conditions, which further can be utilized for exploratory data analysis of dysregulated pathways in disease. These co-expression modules will serve to identify candidate genes for further investigations in tuberculosis disease biology. These resources may aid in the identification of genes with distinct functional roles, and comparative studies such as this can be applied to obtain evidence of condition specific gene expression for further exploration of gene function.

References

- Mayer KH, Dukes Hamilton C. Synergistic pandemics: confronting the global HIV and tuberculosis epidemics. *Clin Infect Dis* 2010;50(Suppl.3):67-70.
- McShane H. Co-infection with HIV and TB: double trouble. *Int J STD AIDS* 2005;16:95-100.
- Manosuthi W, Chottanapand S, Thongyen S, et al. Survival rate and risk factors of mortality among HIV/tuberculosis-coinfected patients with and without antiretroviral therapy. *J Acq Imm Def Syn* 2006;43:42-6.
- Zhang Y. Advances in the treatment of tuberculosis. *Clin Pharmacol Therap* 2007;82:595-600.
- Nayak RR, Kearns M, Spielman RS, Cheung VG. Coexpression network based on natural variation in human gene expression reveals gene interactions and functions. *Genome Res* 2009;19:1953-62.
- Voineagu I, Wang X, Johnston P, et al. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* 2011;474:380-4.
- Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008;9:559.
- Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Gen Mol Biol* 2005;4:17.
- Langfelder P, Luo R, Oldham MC, Horvath S. Is my network module preserved and reproducible? *PLoS Comput Biol* 2011;7:e1001057.
- Dawany N, Showe LC, Kossenkov AV, et al. Identification of a 251 gene expression signature that can accurately detect *M. tuberculosis* in patients with and without HIV co-infection. *PLoS ONE* 2014;9:e89925.
- Davis S, Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* 2007;23:1846-7.
- Database resources of the National Center for Biotechnology Information. *Nucl Acid Res* 2016;44:7-19.
- Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets—update. *Nucl Acid Res* 2013;41:991-5.
- Berry MP, Graham CM, McNab FW, et al. An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature* 2010;466:973-7.
- Ravasz E, Somera AL, Mongru DA, et al. Hierarchical organization of modularity in metabolic networks. *Science* 2002;297:1551-5.
- Yip AM, Horvath S. Gene network interconnectedness and the generalized topological overlap measure. *BMC Bioinformatics* 2007;8:22.
- Langfelder P, Horvath S. Eigengene networks for studying the relationships between co-expression modules. *BMC Syst Biol* 2007;1:54.
- Langfelder P, Zhang B, Horvath S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics* 2008;24:719-20.
- Oldham MC, Horvath S, Geschwind DH. Conservation and evolution of gene coexpression networks in human and chimpanzee brains. *P Natl Acad Sci USA* 2006;103:17973-8.
- Dennis G Jr., Sherman BT, Hosack DA, et al. DAVID: database for annotation, visualization, and integrated discovery. *Gen Biol* 2003;4:P3.
- Miller JA, Oldham MC, Geschwind DH. A systems level analysis of transcriptional changes in Alzheimer's disease and normal aging. *J Neurosci* 2008;28:1410-20.
- Chen M, Gan H, Remold HG. A mechanism of virulence: virulent *Mycobacterium tuberculosis* strain H37Rv, but not attenuated H37Ra, causes significant mitochondrial inner membrane disruption in macrophages leading to necrosis. *J Immunol* 2006;176:3707-16.
- Duan L, Gan H, Golan DE, Remold HG. Critical role of mitochondrial damage in determining outcome of macrophage infection with *Mycobacterium tuberculosis*. *J Immunol* 2002;169:5181-7.
- Perez-Matute P, Perez-Martinez L, Blanco JR, Oteo JA. Role of mitochondria in HIV infection and associated metabolic disorders: focus on nonalcoholic fatty liver disease and lipodystrophy syndrome. *Oxid Med Cell Long* 2013;2013:493413.
- Cossarizza A, Franceschi C, Monti D, et al. Protective effect of N-acetylcysteine in tumor necrosis factor-alpha-induced apoptosis in U937 cells: the role of mitochondria. *Exp Cell Res* 1995;220:232-40.
- Mwandumba HC, Russell DG, Nyirenda MH, et al. *Mycobacterium tuberculosis* resides in nonacidified vacuoles in endocytically competent alveolar macrophages from patients with tuberculosis and HIV infection. *J Immunol* 2004;172:4592-8.
- Zhou H, Xu M, Huang Q, et al. Genome-scale RNAi screen for host factors required for HIV replication. *Cell Host Microb* 2008;4:495-504.
- Zhang Y, Leung DY, Goleva E. Vitamin D enhances glucocorticoid action in human monocytes: involvement of granulocyte-macrophage colony-stimulating factor and mediator complex subunit 14. *J Biol Chem* 2013;288:14544-53.
- Ansari SA, Morse RH. Mechanisms of mediator complex action in transcriptional activation. *Cell Mol Life Sci* 2013;70:2743-56.
- Poss ZC, Ebmeier CC, Taatjes DJ. The mediator complex and transcription regulation. *Crit Rev Biochem Mol Biol* 2013;48:575-608.
- Brass AL, Dykxhoorn DM, Benita Y, et al. Identification of host proteins required for HIV infection through a functional genomic screen. *Science* 2008;319:921-6.
- Ma C, Dong X, Li R, Liu L. A computational study identifies HIV progression-related genes using mRMR and shortest path tracing. *PLoS ONE* 2013;8:e78057.