

RESEARCH ARTICLE

Systematic comparison of differential expression networks in MTB mono-, HIV mono- and MTB/HIV co-infections for drug repurposing

Yao Jiang, Jia-Xuan Zhang, Rong Liu¹*

Hubei Key Laboratory of Agricultural Bioinformatics, College of Informatics, Huazhong Agricultural University, Wuhan, P. R. China

* liurong116@mail.hzau.edu.cn**OPEN ACCESS**

Citation: Jiang Y, Zhang J-X, Liu R (2022) Systematic comparison of differential expression networks in MTB mono-, HIV mono- and MTB/HIV co-infections for drug repurposing. PLoS Comput Biol 18(12): e1010744. <https://doi.org/10.1371/journal.pcbi.1010744>

Editor: Claudio José Struchiner, Fundação Getúlio Vargas: Fundacao Getulio Vargas, BRAZIL

Received: July 5, 2022

Accepted: November 17, 2022

Published: December 19, 2022

Copyright: © 2022 Jiang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data and codes are freely available on GitHub (<https://github.com/hzau-liulab/MTB-HIV>). All other relevant data are within the manuscript and its [Supporting Information](#) files.

Funding: This work was supported by the National Natural Science Foundation of China (32071249 to RL). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

The synergy between human immunodeficiency virus (HIV) and *Mycobacterium tuberculosis* (MTB) could accelerate the deterioration of immunological functions. Previous studies have explored the pathogenic mechanisms of HIV mono-infection (HMI), MTB mono-infection (MMI) and MTB/HIV co-infection (MHCI), but their similarities and specificities remain to be profoundly investigated. We thus designed a computational framework named IDEN to identify gene pairs related to these states, which were then compared from different perspectives. MMI-related genes showed the highest enrichment level on a greater number of chromosomes. Genes shared by more states tended to be more evolutionarily conserved, posttranslationally modified and topologically important. At the expression level, HMI-specific gene pairs yielded higher correlations, while the overlapping pairs involved in MHCI had significantly lower correlations. The correlation changes of common gene pairs showed that MHCI shared more similarities with MMI. Moreover, MMI- and MHCI-related genes were enriched in more identical pathways and biological processes, further illustrating that MTB may play a dominant role in co-infection. Hub genes specific to each state could promote pathogen infections, while those shared by two states could enhance immune responses. Finally, we improved the network proximity measure for drug repurposing by considering the importance of gene pairs, and approximately ten drug candidates were identified for each disease state.

Author summary

Although many efforts have been devoted to exploring the pathogenesis of MTB mono-infection (MMI), HIV mono-infection (HMI) and MTB/HIV co-infection (MHCI), the similarities and differences among the host responses to these infections are not yet illustrated. To address this issue, we constructed the integrated differential expression networks (IDENs) of the three disease states by combining differential and non-differential interactions. We first showed that the IDEN framework could provide a more

Competing interests: The authors have declared that no competing interests exist.

comprehensive picture of host responses than its component methods. Then, we compared the genes in the three IDENs from different viewpoints and found that genes shared by more states showed a stronger preference for certain properties. The correlation analysis and enrichment analysis suggested that MHCI was more similar to MMI rather than HMI, implying that co-infection may be mainly dependent on the pathogenesis of TB. Additionally, we found that hub genes specific to each state may facilitate pathogen infections, whereas hub genes shared by two states may strengthen immune defenses. Finally, we designed gene pair-based network proximity measures to prioritize drug candidates for each disease state, which may provide new insights into the treatment of infections caused by MTB and/or HIV.

Introduction

Tuberculosis (TB) and acquired immune deficiency syndrome (AIDS) are the two most lethal infectious diseases worldwide caused by *Mycobacterium tuberculosis* (MTB) and human immunodeficiency virus (HIV), respectively [1]. The host could be simultaneously infected by MTB and HIV, which promote each other to accelerate the deterioration of immunological functions [2]. HIV infections are the primary risk factor for developing active TB and greatly increase the risk of reactivation of latent TB patients, while MTB infections could accelerate the progression of HIV infections to AIDS and increase the mortality of HIV-infected patients [3,4]. Although intensive efforts have been devoted to investigating the pathogenic mechanisms of the three disease states, namely HIV mono-infection (HMI), MTB mono-infection (MMI) and MTB/HIV co-infection (MHCI) [2], little attention was paid to the similarities and specific differences in the host responses to these states. The increasing transcriptomic data of relevant patients provide unprecedented opportunities to address this problem from a computational perspective. Additionally, the efficacy of existing anti-MTB and anti-HIV drugs becomes less significant, and the specific drugs are lacking for the treatment of co-infections [5]. Therefore, investigation of host responses to these states may provide clues to the repurposing of existing therapeutic agents.

Blood transcriptional profiles could reflect the pathological responses of the host to pathogens and thus have been widely applied to the investigation of MTB or HIV infections. Berry et al. identified 393 transcripts to discriminate between active TB patients and healthy individuals, and these molecules were associated with the type I interferon signaling pathway [6]. Kaforou et al. discovered transcriptional signatures that were converted into disease risk scores to distinguish TB from other conditions in HIV-infected and -uninfected adults and found that these signatures were effective regardless of HIV infections [7]. Devadas et al. compared differentially expressed genes (DEGs) in HIV-1 and HIV-2 infection states. DEGs in HIV-1 infection were involved in critical cellular pathways such as immune cell activation and proliferation, while relatively few DEGs were found in HIV-2 infection [8]. Chen et al. used a meta-analysis strategy to integrate five blood transcriptome datasets and achieve 293 DEGs in co-infection states, four of which could identify HIV patients with or without TB [9]. Based on four TB/HIV co-infection datasets, Duffy et al. used a multi-model machine learning framework to identify 10 transcripts to distinguish disease states associated with co-infection [10]. Generally, these studies extracted gene signatures from expression profiles to identify TB or HIV-related infection states, and the major limitation is that they only focused on individual genes and neglected the relationship between different genes.

Disease progression may be influenced by the rewiring of molecular networks [11]. Thus, integration of blood transcriptional profiles and molecular interactions has been applied to

relevant studies. Sambarey et al. combined RNA sequencing data with protein-protein interactions (PPIs) to develop a sensitive network mining algorithm, which identified a signature comprising 10 genes to distinguish between TB patients and other individuals [12]. In addition, they performed a meta-analysis of multiple TB-related blood transcriptome datasets together with PPI networks to discover 380 common core genes that were highly active in TB disease. Kumar et al. identified 275 differentially expressed host factors that regulated MTB load in human macrophages by siRNA screening, and these molecules were functionally associated through dense interactions [13]. Yoon et al. detected the genetic features of individual samples based on co-expression changes of interacting genes, which can favorably distinguish HIV-1 infection stages and reveal stage-specific signatures [14]. Nevertheless, these studies considered the differential co-expression of gene pairs or the differential expression of genes individually. Combining these two types of information could provide a more comprehensive picture of the host immune response networks. For instance, Sun et al. suggested a differential expression network including differential and non-differential interactions to investigate the dynamics of type 2 diabetes [15]. Sun et al. proposed a consistently differential expression network that merged consistently differential and non-differential interactions across multiple microarray datasets to compare active and latent TB infections [16]. Consequently, the similar methodology could be extended to the comparison of MMI, HMI and MHCI. By investigating the available expression data, however, we found that the existing framework cannot be directly applied to our problem. To facilitate the comparison, we needed to develop a customized algorithm according to the characteristics of relevant data.

Regarding drug repurposing, a variety of network-based algorithms have been developed to prioritize potential drugs for different diseases [17–20]. Especially, the network proximity strategy is receiving increasing attention. Guney et al. first demonstrated that the therapeutic effect of a drug can be inferred from the shortest paths between known drug targets and disease proteins in the human PPI network [21]. Zhou et al. used the same proximity measure to quantify the relationship between drug targets and host proteins associated with four human coronaviruses [22]. Peng et al. developed an improved network proximity metric by considering the weight of target genes and evaluated the proximity between more than 5000 molecular drugs and Alzheimer's disease related genes [23]. However, existing measures only focused on the individual disease proteins and have yet to consider the relationship between drug targets and disease-related protein interactions. Compared to single proteins, as mentioned above, protein interactions may play a more critical role in disease progression. For example, the synergistic effects of amyloid and tau proteins could make a greater impact on Alzheimer's disease than any protein alone [24]. Moreover, the physical interaction between the primary receptor CD4 and the chemokine receptor CCR5 is essential for HIV-1 entry into host cells [25]. Thus, developing the interaction-based proximity metric may offer novel insights into drug repurposing.

In this study, we developed a computational framework named the integrated differential expression network (IDEN), which combined blood microarray data and human protein interactions to identify gene pairs associated with three disease states (i.e. MMI, HMI and MHCI). First, we constructed the IDEN for each disease state and evaluated the confidence of derived results. Then, we illustrated the similarities and specificities of genes and gene pairs in the three networks from different aspects. Especially, we explored the role of hub genes in sub-networks composed of gene pairs related to different infection states. Finally, we designed network proximity measures by considering the importance of gene pairs and used the metrics to prioritize reusable drugs against the three types of infections. The present work not only improves our understanding of MTB- and/or HIV-infection but also provides novel clues to

the treatment of patients with relevant diseases. The data and codes are freely available on GitHub (<https://github.com/hzau-liulab/MTB-HIV>).

Materials and methods

Dataset

Blood microarray data. In this work, the gene expression profile datasets were obtained by searching the Gene Expression Omnibus (GEO) database [26]. We adopted ‘tuberculosis’ and ‘HIV’ as the keywords and selected ‘*Homo sapiens*’ and ‘Expression profiling by array’ as the species and study type, respectively. To reduce the influence of batch effects, we only retained the datasets composed of whole blood samples from the GPL10558 platform, because the data of this platform were sufficient to perform meaningful analyses. As shown in S1 Table, we collected four datasets for each disease state (i.e. MMI, HMI and MHCI). Note that the latent TB patients were not included in these datasets. To ensure the similar numerical range of expression profiles, the relevant datasets were processed by log2 transformation and quantile normalization using the *limma* package (S1 Fig) [27]. The probe IDs were converted into gene symbols based on the annotation file, and probes with missing values were excluded from further analyses. When multiple probes corresponded to the same gene, the average value was calculated to represent the gene expression level.

Human PPI network. Human PPIs were extracted from the HIPPIE database (v2.2), which collects a total of 411430 PPIs between 18166 proteins from several well-established PPI databases [28]. Self-interactions and duplicates were removed and only PPIs with confidence scores greater than 0 were reserved. Finally, the human PPI network included 400423 unique PPIs comprising 18141 proteins.

Drug–target interaction network. The drug–target interactions were extracted from the DrugBank database (v5.17) and the Therapeutic Target Database (TTD, June 2020) [29,30]. We obtained 10770/2506 interactions between 2079/2115 approved drugs and 2575/721 targets from DrugBank/TTD. We merged the data from two databases and only retained the targets present in the human PPIs. A drug–target network having 10061 interactions between 2065 drugs and 2159 targets was constructed in this work.

Functionally important gene sets. We collected 212 HIV-related genes involved in the ‘Human immunodeficiency virus 1 infection’ pathway (hsa05170) and 180 TB-related genes involved in the ‘Tuberculosis’ pathway (hsa05152) from the KEGG database. Furthermore, 2828 inflammatory genes were obtained by searching the keyword ‘inflammatory’ in the gene database of NCBI; 1575 human essential genes were extracted from the study of Hart et al. [31]; and 3659 housekeeping genes that were uniformly expressed in all human tissues were derived from the study of Eisenberg and Levanon [32]. We also downloaded virus–human, bacteria–human and fungi–human PPIs from pathogen–host interaction databases (e.g. PATRIC (November 2016), HPIDB, VirHostNet (January 2019) and PHISTO (September 2019)) [33–36]. We achieved 5695, 3394 and 609 targets for viruses, bacteria and fungi, respectively.

Differential co-expression analysis (DCA)

As shown in Fig 1, we evaluated the expression correlation between genes using the Pearson correlation coefficient (PCC). The PCC values of each gene pair in the disease (i.e. MMI, HMI or MHCI) and normal states were calculated. Since the samples associated with each disease state were derived from four datasets (S1 Table), we could obtain four PCC values for each gene pair using the individual datasets. The mean PCC value was calculated as the representative. Similarly, we computed the average PCC value of each gene pair in the healthy group

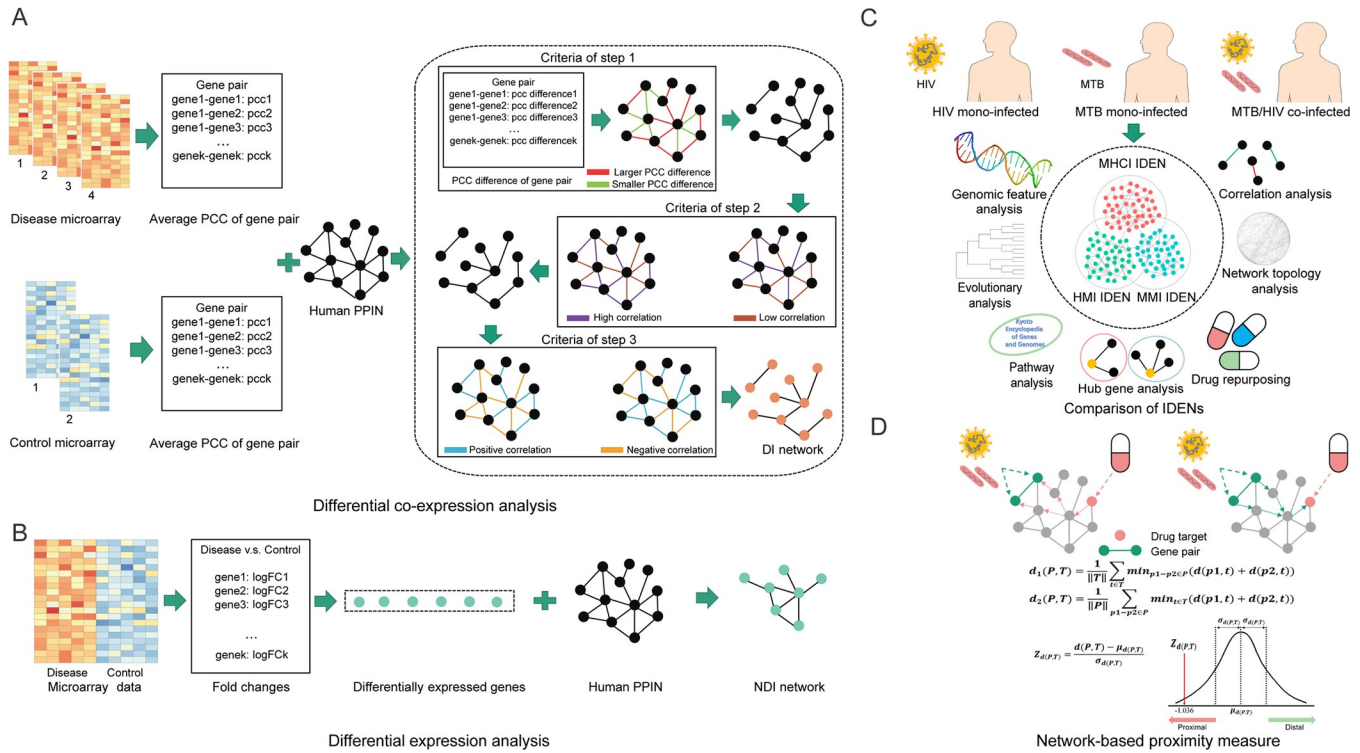


Fig 1. IDEN algorithm and gene pair-based drug repurposing. (A) Flowchart of the DCA method. (B) Flowchart of the DEA method. (C) Comparison of three IDENs. (D) Gene pair-based drug repurposing.

<https://doi.org/10.1371/journal.pcbi.1010744.g001>

using the GSE29429 and GSE83456 datasets. The resulting PCC values could be mapped to the corresponding PPIs, which generated the weighted PPI networks for the disease and normal states, respectively. Then, we computed the difference in PCC values (abbreviated as PCC difference) between the two states for gene pairs and designed a three-step method to extract the gene pairs having significant differential changes. First, we borrowed the idea of the three-sigma rule to identify gene pairs with larger absolute differences. A gene pair was selected if its absolute difference (i.e. |PCC difference|) was greater than the mean plus three standard deviations of |PCC difference| of all gene pairs. Second, we evaluated whether the gene pairs exhibited a difference in the correlation status (high or low correlation). A gene pair was selected if its absolute value of PCC (i.e. |PCC|) was greater than 0.5 (i.e. an empirically assigned cutoff) in one state and less than 0.5 in the other state. Third, we examined whether the gene pairs showed a difference in the sign of correlation values (positive or negative correlation). A gene pair was selected if its PCC was greater than 0 in one state and less than 0 in the other state. The process of three steps is shown in S2 Fig, where the finally reserved gene pairs were defined as differential interactions (DIs). As a control, we randomly permuted the expression values of genes and repeated this procedure 100 times. The one-sample *t*-test was used to evaluate the significance of DIs by comparing the native |PCC difference| with the mean of |PCC difference| derived from random permutations. The results suggested that the *p*-values of all DIs were significant (*p* < 0.05).

Differential expression analysis (DEA)

In addition to DIs, we identified non-differential interactions (NDIs) involved in various disease states. The NDI generally had trivial changes in expression correlation, but this

interaction may also be associated with disease progression if its two genes were DEGs. Because GSE37250, GSE39939 and GSE39940 contained only disease samples and had a clearly different numerical range relative to the normal samples of GSE29429 and GSE83456, these three datasets were excluded from further analyses. In other words, only one dataset could be used for each disease state. For GSE29429 and GSE83456, the DEA procedure was performed directly because they included normal samples. However, there were no normal samples in GSE69581. Considering that GSE69581 and GSE83456 shared a similar numerical range, we combined the samples from these two datasets and removed the batch effect using the ComBat program [37]. Then, the normal samples of GSE83456 could be considered as the control data of GSE69581 (S3 Fig). Based on each dataset, we computed the absolute value of the log₂ fold change (i.e. $|\log_2FC|$) for each gene. Due to the different distribution of $|\log_2FC|$ for the three datasets, a unified threshold may not be suitable for detecting DEGs (S4 Fig), so we continued to adopt the mean plus three standard deviations of $|\log_2FC|$ as the adaptive threshold. Genes with $|\log_2FC|$ greater than this cutoff and $FDR < 0.05$ were treated as DEGs. Subsequently, DEGs were mapped to the PPI network and a gene pair was defined as a NDI if both genes were DEGs.

Genomic characteristics analysis

We checked the chromosomal distribution of disease-related genes by calculating the ratio of the number of identified genes on each chromosome to the total number of identified genes. Furthermore, we extracted single nucleotide polymorphisms (SNPs) located in human protein-coding genes from the Ensembl database (release 104) [38]. The SNP density of each gene was defined as the ratio of the number of mapped SNPs to the gene length. Disease-related SNPs (dSNPs) were obtained from the ClinVar database (November 2019) [39]. We computed the proportion of dSNPs among total SNPs in the detected genes. Based on gene-disease associations from the DisGeNET database (v7.0), we assessed the proportion of disease-causing genes among our identified genes [40]. The distances between genes on the same chromosome were also calculated in this work.

Evolutionary and expression analysis

Here, we investigated three evolutionary characteristics of disease-related genes, including the dN/dS ratio, protein age and homologous gene number. Specifically, human-mouse orthologous genes were derived from the Ensembl BioMart database, and the dN/dS ratio was calculated to illustrate the gene evolutionary rate [38]. The protein age was extracted from the ProteinHistorian database to infer the evolutionary origin of proteins [41]. The homologous gene number was computed based on the HomoloGene database (build68) [42]. Known sites of post-translational modifications (PTMs) in human proteins were derived from the PhosphoSitePlus database (v6.6) [43]. The proportion of modification sites in each protein was calculated to evaluate the PTM level of identified genes.

Network topology analysis

Based on the human PPI network, we investigated the topological features of identified genes, including the degree centrality, closeness centrality, betweenness centrality and shortest path distribution. The degree centrality is the number of genes that are directly connected to the target gene. The closeness centrality is defined as the inverse of the average shortest path distance from the target gene to other genes in the network. The betweenness centrality is defined as the sum of the fraction of shortest paths between all pairs of genes that pass through the target gene. The shortest path between two genes could be calculated using the Dijkstra

algorithm. Furthermore, we adopted the Markov clustering (MCL) algorithm with default parameters to extract topological modules in the PPI network [44]. To evaluate the topological distribution of identified gene pairs, we computed the intramodule to intermodule interaction ratio ($MIR = N_{intra\text{module}}/N_{inter\text{module}}$), where $N_{intra\text{module}}$ denotes the number of interactions between two genes from the same module, and $N_{inter\text{module}}$ denotes the number of interactions between two genes from different modules.

Network proximity measure for drug repurposing

Existing drug repurposing algorithms generally used network-based measures to prioritize potential drugs according to the proximity between known drug targets and disease-related genes in the human PPI network [21]. Considering the importance of gene pairs found in this work, we proposed an alternative strategy to evaluate the effectiveness of drugs by calculating the shortest distance between drug targets (T) and disease-related gene pairs (P). Here, three interaction-based proximity measures could be designed for each drug as follows:

$$d_1(P, T) = \frac{1}{\|T\|} \sum_{t \in T} \min_{p1-p2 \in P} (d(p1, t) + d(p2, t))$$

$$d_2(P, T) = \frac{1}{\|P\|} \sum_{p1-p2 \in P} \min_{t \in T} (d(p1, t) + d(p2, t))$$

$$d_3(P, T) = d_1(P, T) + d_2(P, T)$$

where $d(p1, t)$ and $d(p2, t)$ denote the shortest distance from genes $p1$ and $p2$ to drug target t in the network. The first distance measure is to fix the drug target set and compute the shortest distance to the elements in the PPI set, while the second measure is to fix the PPI set and compute the shortest distance to the elements in the target set. Meanwhile, we constructed a reference distribution for each drug. Referring to the known drug targets, we randomly selected the same number of proteins with similar degree distributions from the whole PPI network. Due to the scale-free nature, a small portion of proteins have extremely high degrees. To avoid the repetitive selection of these proteins, a binning approach was proposed to divide proteins into different intervals such that at least 100 proteins were included in each bin. We repeated the permutation process 1000 times and the original measures were converted into Z-scores as follows:

$$Z_{d(P,T)} = \frac{d(P, T) - \mu_{d(P,T)}}{\sigma_{d(P,T)}}$$

where $\mu_{d(P,T)}$ and $\sigma_{d(P,T)}$ represent the mean and standard deviation of the results from 1000 permutation tests, respectively. According to the normal distribution, a drug was considered to be potentially effective for a given disease if $Z \leq -1.036$ (i.e. a Z-score that was lower than 85% of the reference scores was considered significant in the one-sided test) [45,46].

Connectivity Map (CMAP) based drug repurposing

Besides the network-based approach, the GSEA program in CMAP was adopted to evaluate the potential of drug candidates by comparing the gene expression profiles upon drug treatment with the gene expression profiles of patients [47]. Because the maximum allowed number of submitted genes is 150, we ranked DEGs based on the $|\log_2FC|$ measure for each disease state and used the top 150 up- and down-regulated DEGs as the gene signature. If the GSEA score is negative (i.e. a negative correlation between the drug and signatures), the drug could

reverse the disease. Thus, the GSEA score can be used as an auxiliary metric for assessing the selected drugs. Note that not all drugs can retrieve a score from CMAP.

Statistical analysis

Hypergeometric tests were adopted to estimate the significance of the overlap of gene sets. For a given attribute, the differences between the two groups of genes or gene pairs were analyzed using Wilcoxon rank sum tests. In this work, $p < 0.05$ was considered statistically significant. All statistical analyses were performed using the R package.

Results

Overview of three differential expression networks

As shown in [Fig 1A and 1B](#), we integrated microarray data and protein interactions to identify gene pairs involved in different pathogen infections. By taking the union of the results of DCA and DEA methods, we obtained the IDEN of MMI comprising 911 gene pairs and 1007 genes (termed MMI-P and MMI-G, respectively), the IDEN of HMI comprising 1314 gene pairs and 1266 genes (termed HMI-P and HMI-G, respectively) and the IDEN of MHCI comprising 1363 gene pairs and 1576 genes (termed MHCI-P and MHCI-G, respectively). We found that there was a significant overlap between any two IDENs ([Fig 2A and 2B](#)), suggesting that the three disease states could be related and share common pathogenic mechanisms. Particularly, the overlaps between the MHCI-related and MMI-/HMI-related IDENs were more remarkable than those between the MMI-related and HMI-related IDENs. This indicated that the co-infected state could be more closely associated with two mono-infected states. The DCA method found 604, 851 and 1290 DIs for MMI, HMI and MHCI, respectively, which were the major components of IDENs ([Fig 2C](#)). The identical DIs between the two mono-infected states were clearly less than those between the mono-infected and co-infected states, which can explain the above observation for the whole IDENs. The DEA method identified 307, 477 and 74 NDIs for three states ([Fig 2E](#)). The identical NDIs between MMI and MHCI accounted for 83.8% (62/74) of the total NDIs of MHCI, while those between HMI and MHCI only accounted for 5.4% (4/74). This result indicated that MHCI would be more similar to MMI from the viewpoint of gene expression changes, implying that TB disease plays a dominant role in co-infection. We also compared the results inferred by the DCA and DEA methods and found that there were small overlaps ([Fig 2G–2L](#)). Thus, the IDEN that was a combination of DIs and NDIs may provide more comprehensive insights into host immune responses.

To facilitate the comparison and analysis, the gained genes and gene pairs were separated into different categories. All the genes and gene pairs associated with the three disease states were termed IDEN-G and IDEN-P, respectively ([S5 Fig](#)). We divided IDEN-G into seven gene groups, including HMI-specific genes (HMI-SG), MMI-specific genes (MMI-SG), MHCI-specific genes (MHCI-SG), shared genes between HMI and MMI (HMI-MMI-SG), shared genes between HMI and MHCI (HMI-MHCI-SG), shared genes between MMI and MHCI (MMI-MHCI-SG), and common genes among the three disease states (Common-G). The similar partition was performed for IDEN-P. After excluding IDEN-G and IDEN-P from the human PPI network, the remaining genes and gene pairs were considered the control genes and gene pairs, namely Control-G and Control-P.

Advantage and confidence of the IDEN framework

To evaluate the biological relevance of the identified genes, we collected genes associated with HIV or MTB infections from existing computational and experimental resources. There were

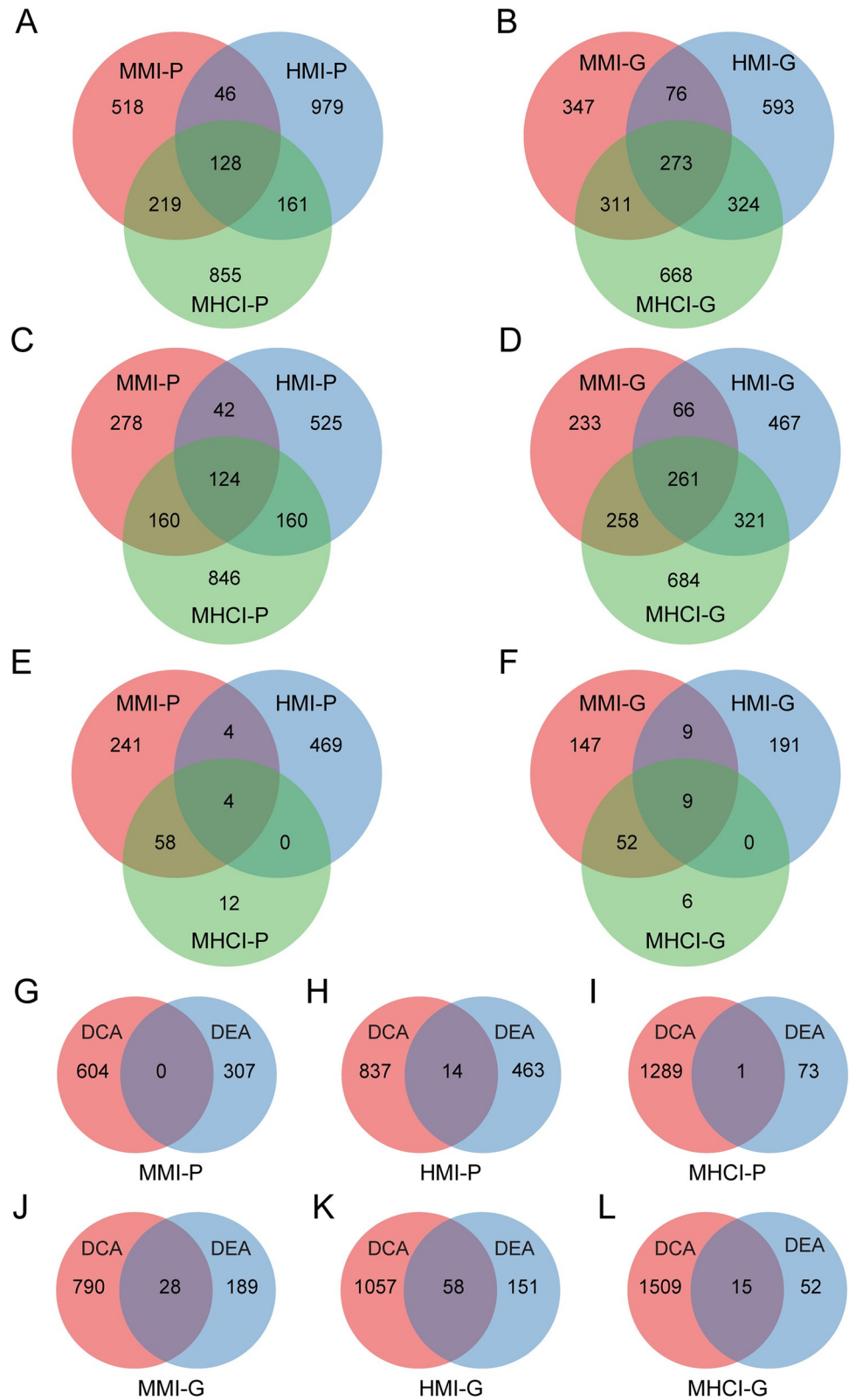


Fig 2. Overview of MMI-, HMI- and MHCI-related gene pairs and genes. (A-B) Number of gene pairs and genes identified by the IDEN framework. (C-D) Number of gene pairs and genes identified by the DCA method. (E-F) Number of gene pairs and genes identified by the DEA method. (G-I) Comparison of gene pairs identified by the DCA and DEA methods. (J-L) Comparison of genes identified by the DCA and DEA methods.

<https://doi.org/10.1371/journal.pcbi.1010744.g002>

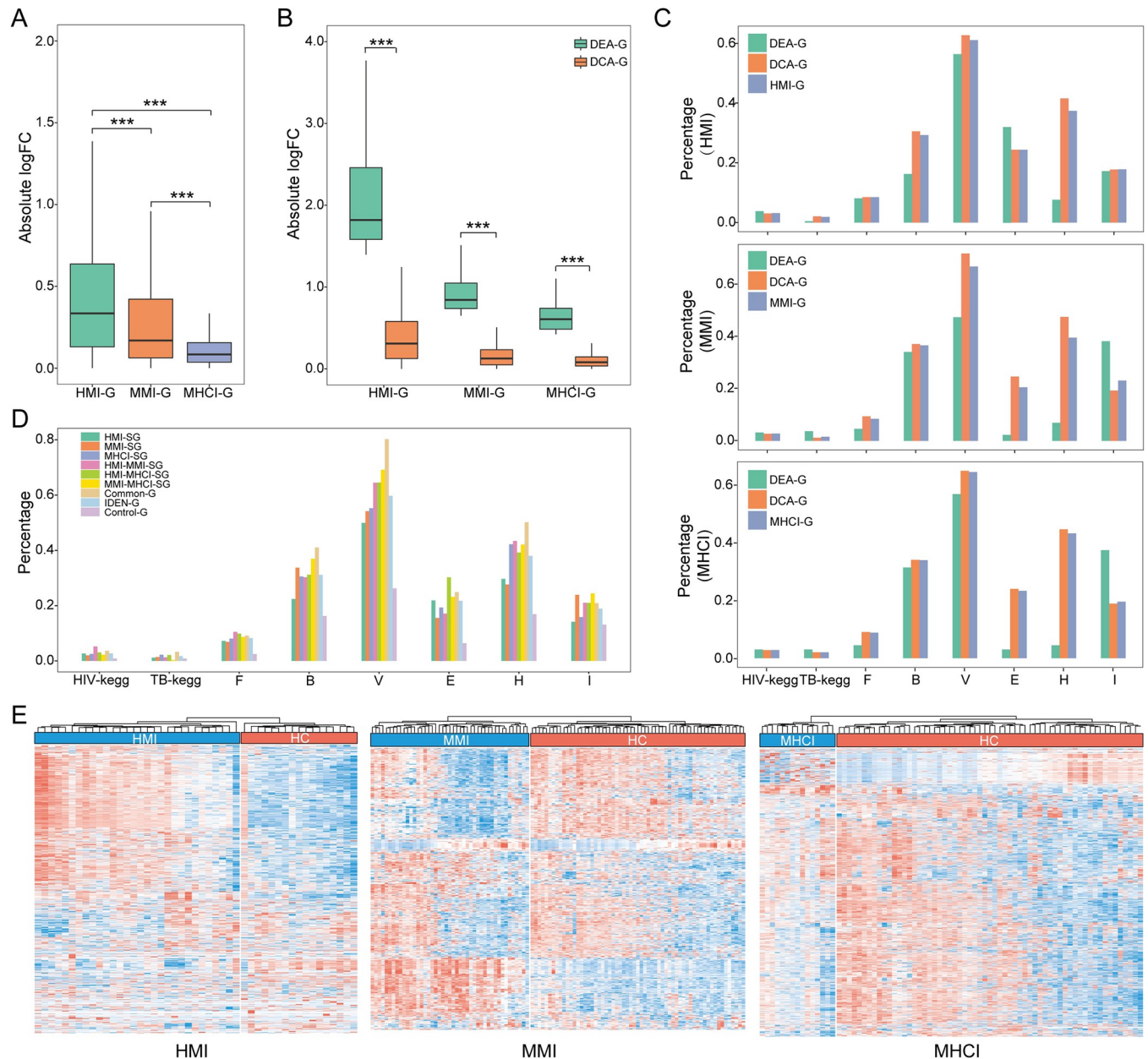


Fig 3. Advantage and confidence of IDEN framework. (A) Expression change of identified genes. (B) Expression change of genes identified by the DCA and DEA methods. (C) Enrichment of genes identified by the DCA and DEA methods in the important gene sets. F, B and V denote human targets for fungi, bacteria and viruses, respectively, and E, H and I denote essential genes, housekeeping genes and inflammatory genes, respectively. (D) Enrichment of the subclasses of IDEN-G in the important gene sets. (E) Clustering analysis of IDEN-G in the important gene sets. The rows denote disease-related genes (red: up-regulated and blue: down-regulated) and the columns denote samples. *** $P < 0.001$, ** $0.001 \leq P < 0.01$ and * $0.01 \leq P < 0.05$.

<https://doi.org/10.1371/journal.pcbi.1010744.g003>

significant overlaps between the three group of genes (i.e. MMI-G, HMI-G and MHCI-G) and the available resources (S2 Table), indicating that the inferred genes were highly involved in the relevant diseases. Then, we compared the expression changes of these genes between disease and normal states (Fig 3A and 3B). HMI-G showed the greatest changes, followed by MMI-G, and MHCI-G showed the smallest changes. During the co-infected state, the expression levels of disease-related genes had trivial changes, but their expression correlations may

change dramatically. For each status, genes identified by the DEA (termed DEA-G) had greater changes, whereas those identified by the DCA (termed DCA-G) had relatively mild changes ($|\log_2FC| < 0.5$). Thus, the DCA (which depends on expression correlation changes) could complement the shortcoming of the DEA (which focuses on DEGs).

Moreover, we profoundly compared the results derived from the two component methods. [S3 Table](#) displays the enrichment of DCA-G and DEA-G in the seven gene sets including different functional genes, namely HIV- and TB-related genes in KEGG pathways, human target genes of various pathogens (e.g. fungi, bacteria and viruses) and other important host genes (e.g. essential genes, housekeeping genes and inflammatory genes). Compared with DEA-G, DCA-G was generally more enriched in these gene sets ([Fig 3C](#)), indicating that the DCA has advantages in identifying potential disease genes associated with pathogen infections. Furthermore, the IDEN identified more valuable genes than the DCA in all gene sets ([S3 Table](#)), suggesting that the DEA can also complement the DCA and was indispensable for the proposed framework. Additionally, we explored the relationship between these gene sets and the different subclasses of IDEN-G ([S4 Table](#)). IDEN-G and its subclasses showed consistently higher enrichments than Control-G ([Fig 3D](#)). Common-G yielded the highest measure for four gene sets, including TB pathway genes, human target genes of bacteria and viruses and housekeeping genes, indicating that Common-G may play a key role in the three disease states.

To further evaluate the confidence of our results, the identified genes and gene pairs were adopted as signatures to distinguish between disease and control samples. Specifically, the expression values of MMI-G, HMI-G and MHCI-G were utilized as features to construct the random forest models using the Scikit-learn package. All parameters except the number of trees ($n_estimators = 500$) were default values. We evaluated our classifiers using the leave-one-out cross-validation. All the three groups of genes showed favorable performances, and MHCI-G achieved the greatest area under the curve ($AUC = 0.993$, [S5 Table](#)). In addition, we checked the prediction capabilities of MMI-P, HMI-P and MHCI-P by implementing Zhang et al.'s algorithm [48]. Generally, the performances of gene pair-based models were comparable to the results of gene-based models, and MHCI-P yielded the best performance ($AUC = 1$, [S5 Table](#)). We also compared the IDEN signatures with seven gene signatures in previous studies. Our signatures performed more favorably on the identification of MHCI than other signatures and achieved slightly better or comparable performance on the prediction of MMI and HMI ([S6 Table](#)). We then performed the clustering analysis based on the gene signatures. Compared with the healthy samples, the disease samples displayed clearly different expression patterns ([Fig 3E](#)), again demonstrating the reliability of our results.

Genomic characteristics analysis

To characterize the achieved genes at the genomic level, we first explored their chromosomal distributions. The analysis illustrated that these disease-related genes were not restricted within specific chromosomes but widely distributed in the human genome. Approximately 97% of these genes were located on autosomes ([S6A Fig](#)). Moreover, MMI-G, HMI-G and MHCI-G demonstrated the highest enrichment on 8, 5 and 10 chromosomes, respectively ([S1 Data](#)). SNPs are the most prevalent genetic variants in the human genome. We computed the SNP density of genes in the three groups. The SNP density of MMI-G was higher than the measure of the other two gene groups on 9 chromosomes, and MHCI-G and HMI-G displayed the highest density on 7 and 7 chromosomes, respectively ([S6B Fig](#)). The distribution of disease SNPs showed that MMI-G, HMI-G and MHCI-G achieved the highest proportion on 11, 7 and 5 chromosomes, respectively ([S6C Fig](#)). Also, we investigated the involvement of three gene sets in human diseases using known disease-causing genes as the reference. The

proportion of disease-causing genes among our genes on each chromosome was more than 50%. MMI-G obtained a higher proportion on 11 chromosomes than the other two groups (S6D Fig). We also checked the relative positions of gained genes in the genome. The distances of genes from the seven subsets of IDEN-G were significantly different (S7 Fig). In particular, MMI-SG had the closest genomic distance, implying that they tended to be clustered together on the same chromosome.

Evolutionary and expression analysis

In this section, we investigated the evolutionary properties of genes detected by the IDEN framework. IDEN-G possessed lower dN/dS ratios than Control-G and thus showed a stronger purifying selection ability (Fig 4A). Moreover, IDEN-G held older protein ages and higher homologous gene numbers (Fig 4B and 4C). These results were consistent with previous findings that pathogens tended to target relatively conserved genes in the host [49]. Among the seven subclasses, Common-G showed the strongest evolutionary measures, implying that common genes may play indispensable roles in the three infection statuses. And genes shared by two statuses (i.e. HMI-MMI-SG, HMI-MHCI-SG and MMI-MHCI-SG) showed more conserved measures than genes specific to each status (i.e. HMI-SG, MMI-SG and MHCI-SG). Thus, genes with a greater overlapping degree could be more functionally important from an evolutionary viewpoint. Regarding the PTM levels, the protein products of IDEN-G had a higher proportion of phosphorylated, ubiquitinated, acetylated and methylated residues (Fig 4D–4G), implying that disease-related genes probably disrupt cellular activities through different PTMs. Actually, modifications of the host proteome were often manipulated by pathogens to influence the outcome of infections [50]. Additionally, we observed that genes shared by more statuses had more PTM sites.

Based on the two datasets from different references (GSE37250 and GSE39940), the average expression value of all samples in each dataset was computed to represent the expression level of each gene. IDEN-G was more abundantly expressed compared to Control-G (Fig 4H and 4I). Additionally, Common-G held the highest expression level among all subclasses, followed by MMI-MHCI-SG, while HMI-SG had the lowest level. The differences among the other subclasses were not so significant. We divided all samples into different groups (i.e. HMI, MMI and MHCI) and observed similar expression patterns (S8 Fig), implying that the mechanisms of gene expression may be shared by the three states.

Network topology analysis

To understand the network context of disease-related genes, we computed three topological features, including degree centrality, closeness centrality and betweenness centrality. The degree of IDEN-G was significantly greater than that of Control-G (Fig 5A), suggesting that pathogens may manipulate relevant pathways by targeting highly connected genes. Compared with Control-G, IDEN-G had higher closeness measures, indicating that these genes could be close to the center of the PPI network and easily transmit signals to other relevant genes (Fig 5B). The greater betweenness measures suggested that IDEN-G was located on the shortest paths and can therefore control the information flow among genes (Fig 5C). We also found that genes shared by more statuses had an increasing value for these topological features, illustrating that these genes occupied more important positions in the network.

We then computed the distribution of shortest paths between genes. Compared to Control-G, IDEN-G generally had shorter shortest paths, suggesting that these genes tended to co-localize and form modules in the PPI network (Fig 5D). Especially, this tendency was more remarkable for the genes with a greater overlapping degree (e.g. Common-G and

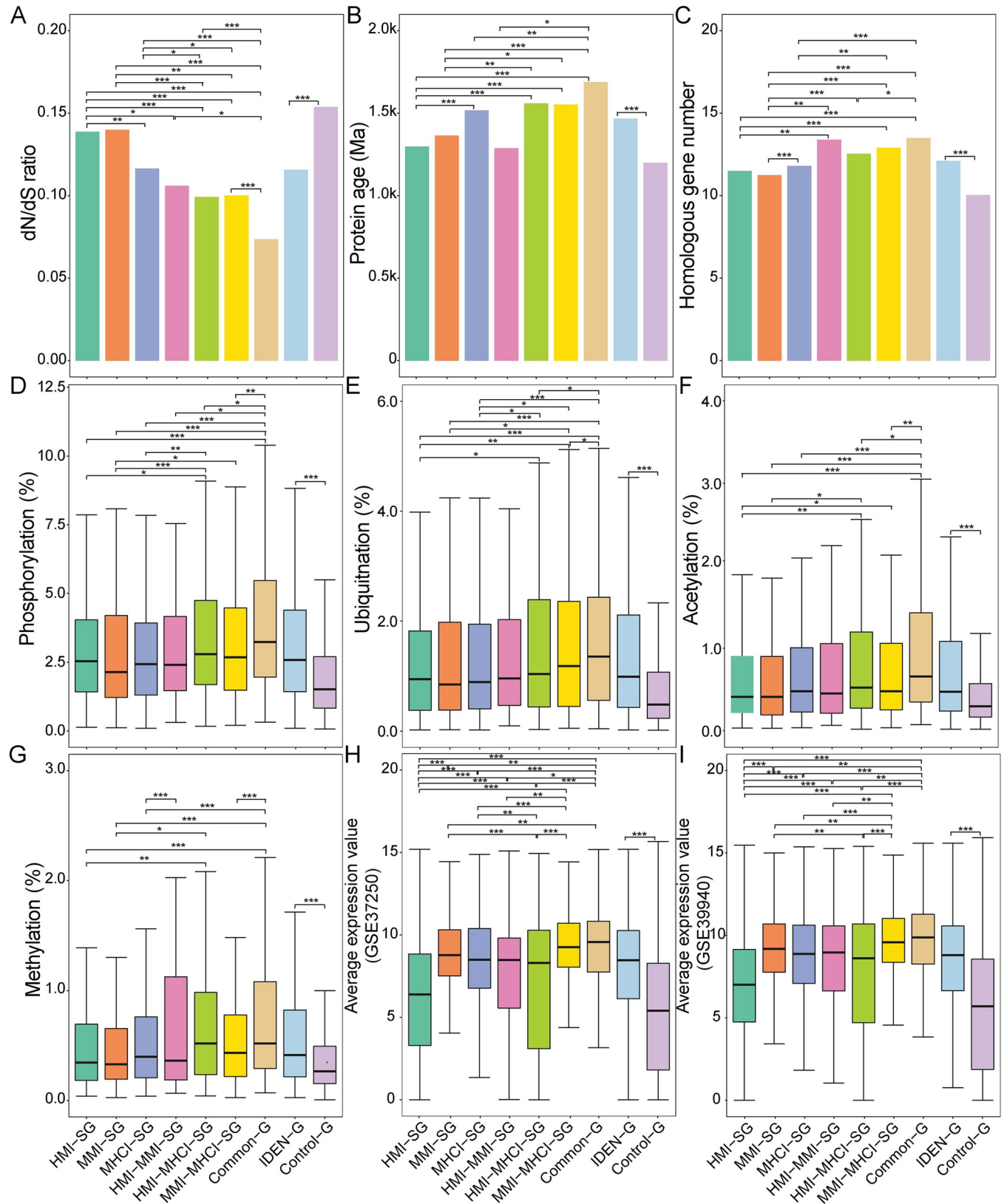


Fig 4. Evolutionary and expression features of disease-related genes. (A) dN/dS ratio. (B) Protein age. (C) Number of homologous genes. (D) Proportion of phosphorylated sites. (E) Proportion of ubiquitinated sites. (F) Proportion of acetylated sites. (G) Proportion of methylated sites. (H) Expression level of detected genes (GSE37250). (I) Expression level of detected genes (GSE39940). *** $P < 0.001$, ** $0.001 \leq P < 0.01$ and * $0.01 \leq P < 0.05$.

<https://doi.org/10.1371/journal.pcbi.1010744.g004>

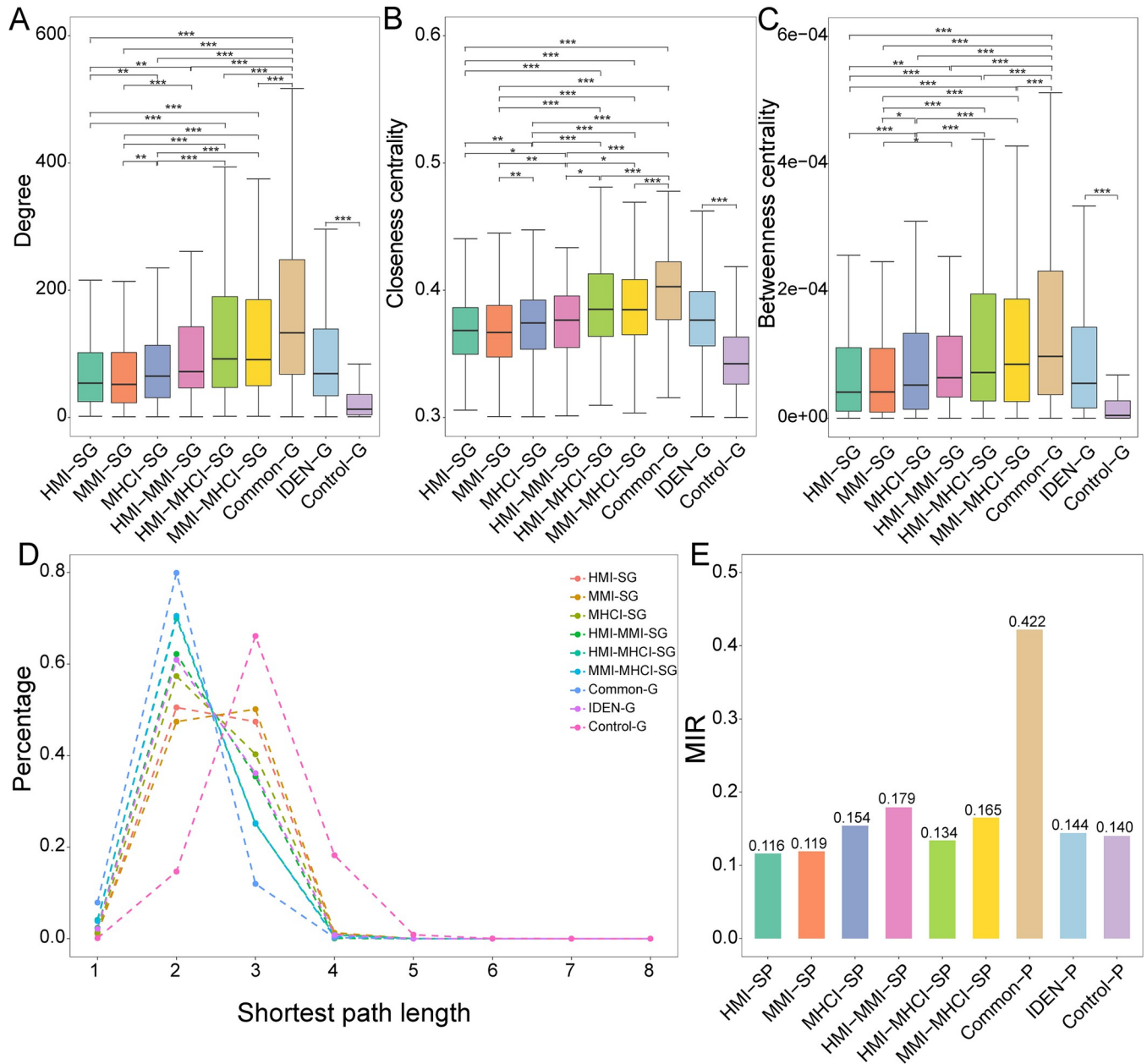


Fig 5. Network topology features of disease-related genes. (A) Degree centrality. (B) Closeness centrality. (C) Betweenness centrality. (D) Shortest path length. (E) MIR measure. *** $P < 0.001$, ** $0.001 \leq P < 0.01$ and * $0.01 \leq P < 0.05$.

<https://doi.org/10.1371/journal.pcbi.1010744.g005>

HMI-MHCI-SG). Besides, the intramodule to intermodule interaction ratio was computed to evaluate the topological distribution of gene pairs. We obtained 601 modules consisting of 12518 genes by the MCL algorithm [44]. The MIR of common pairs (Common-P) reached 0.422 (Fig 5E), indicating that the relevant genes were tightly clustered within topological modules, possibly due to their critical roles in the pathogenesis. MHCI-SP, HMI-MMI-SP and MMI-MHCI-SP also achieved relatively higher MIR values than Control-P. Accordingly, the overlapping genes and gene pairs tended to be more adjacent in the network.

Expression correlation analysis

This section focuses on the comparison of gene pairs from the expression level. As shown in Fig 6A, all the three groups of specific gene pairs (i.e. HMI-SP, MMI-SP, and MHCI-SP) obtained higher |PCC| values in the disease group than in the healthy group, indicating that

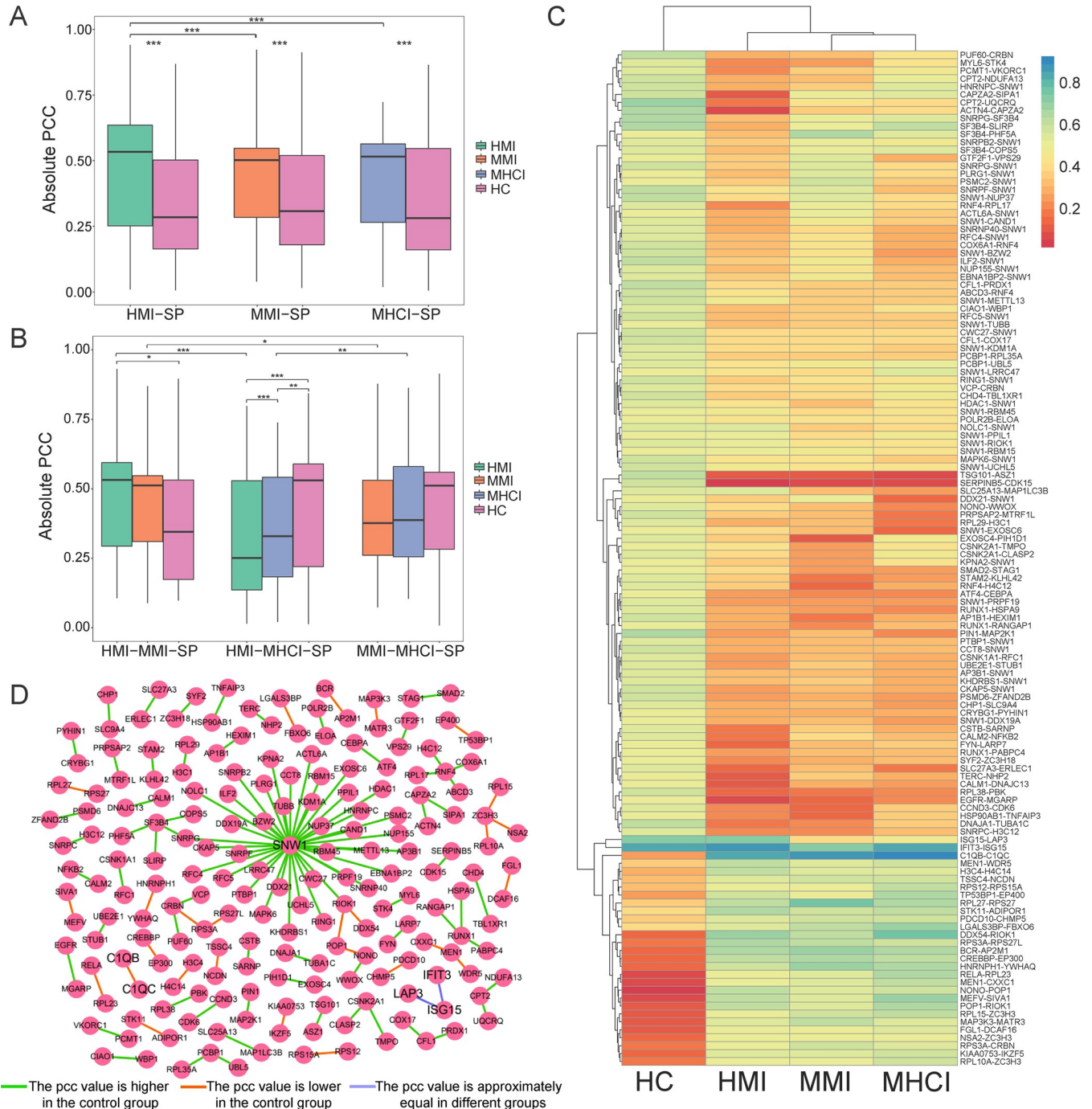


Fig 6. Expression correlations of disease-related gene pairs. (A) Absolute PCC values of specific gene pairs. (B) Absolute PCC values of gene pairs shared by two statuses. (C) Absolute PCC values of common gene pairs. (D) Network of common gene pairs. If the absolute PCC values are consistently greater than (or less than) 0.5 in different groups, the expression correlation of a gene pair is considered to be approximately equal. *** $P < 0.001$, ** $0.001 \leq P < 0.01$ and * $0.01 \leq P < 0.05$.

<https://doi.org/10.1371/journal.pcbi.1010744.g006>

the associated genes were more strongly interconnected during the specific infection stage. The |PCC| value of HMI-SP was significantly higher than that of the other specific groups. This implied that the protein interactions specific to HMI were more activated. Moreover, we found that HMI-MMI-SP had higher |PCC| values in the two disease states (Fig 6B). However, the opposite trend was observed for HMI-MHCI-SP and MMI-MHCI-SP, suggesting that the correlations of shared pairs related to co-infections decreased significantly in the disease states.

Based on the expression correlations of common pairs, we found that MMI and MHCI shared more similarities (Fig 6C). Among Common-P, 27 gene pairs showed high correlations in the disease state and low correlations in the normal state, while 99 gene pairs had an opposite trend. The *CIQA*, *CIQB* and *CIQC* collectively constitute the complement *CIQ*, which is essential to the activation of the classical complement pathway. The correlation of *CIQB-CIQC* was higher and gradually increased in the three disease states (MHCI > MMI > HMI), illustrating that the interactions between complement components were enhanced for MTB-infected individuals, especially the co-infected patients. Interestingly, 40 pairs with lower correlations in the disease state included a multifunctional protein *SNWI* (Fig 6D), which acts by protein interactions, mRNA splicing regulation and transcriptional control [51]. Accordingly, this protein may be the primary host factor targeted by pathogens and the associated interactions would weaken or even disappear in the disease state.

Comparative analysis of functional pathways

The KEGG and GO enrichment analysis was performed using the clusterProfiler package [52]. We retained the top 20 pathways with FDR < 0.05. As shown in Fig 7, the three disease states shared seven pathways, such as 'ribosome', 'spliceosome' and 'coronavirus disease-COVID-19'. The use of ribosomal pathway was possibly because the host needs to synthesize a large number of proteins in response to pathogen infections [16]. The spliceosome mediates pathogen infections by participating in mRNA splicing, which enables pathogens to prevent the downstream immune responses [53]. Besides, MMI and MHCI shared several additional pathways, such as 'proteasome' and 'prion disease'. The proteasome regulates the degradation of most cellular proteins and removes erroneous proteins [54]. MTB may thus cause immune dysregulation by disturbing the function of proteasomes. In contrast, HMI shared only one pathway with MMI or MHCI. The GO analysis confirmed that the inferred genes were enriched in RNA-related biological processes, and MMI shared more processes with MHCI than HMI (S9 Fig). Overall, MMI was more similar to MHCI according to functional analyses, implying that the pathogenesis of co-infection may be dominated by MTB [55].

Nuclear factor-kappa B (NF- κ B) was indispensable for the progression of MTB infection and significantly associated with the three disease states (MMI: FDR = 0.003, HMI: FDR = 0.006 and MHCI: FDR = 0.0002), so we profoundly explored this pathway [56]. As shown in Fig 8A, our framework identified 41 genes. Compared to MMI-SG and HMI-SG, MHCI-SG was more widely distributed throughout the pathway and may control the degradation of the suppressor protein I κ B. Furthermore, genes in the *TRAF* family (*TRAF2*, *TRAF5* and *TRAF6*) were included in both HMI-SG and HMI-MHCI-SG, implying that these genes may be involved in HIV infection. In fact, the three proteins could interact with HIV-1 Nef proteins to initiate NF- κ B activation and increase HIV-1 replication [57]. The p50 and RelB proteins are involved in the formation of NF- κ B dimers in the classical and non-classical pathways, respectively. The former was shared by HMI and MHCI, and the latter was shared by MMI and MHCI. This result implied that the p65-p50 dimer may be formed for HIV-infected patients, while the p52-RelB dimer may be generated for MTB-infected patients. Common-G was generally located in downstream positions of the pathway and included p65 and p100,

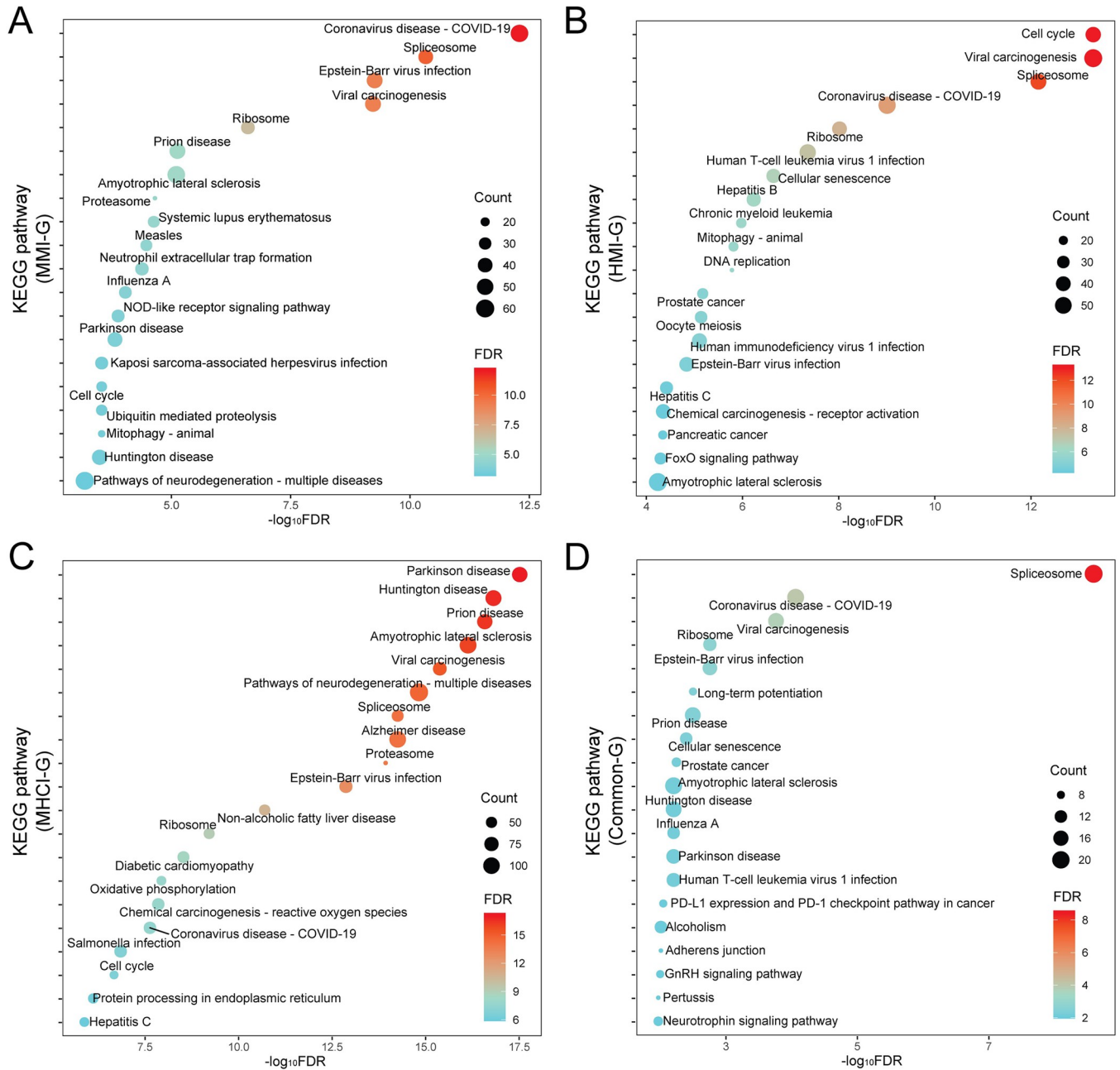


Fig 7. KEGG pathway analysis of disease-related genes. (A) MMI-G. (B) HMI-G. (C) MHCI-G. (D) Common-G.

<https://doi.org/10.1371/journal.pcbi.1010744.g007>

which are also essential for NF- κ B dimerization. In Fig 8B, we visualized PPIs related to this pathway. A greater number of MMI-specific pairs (MMI-SP) were observed, further confirming the importance of this pathway in MTB infection.

Analysis of hub genes in different sub-networks

In this work, we defined the top five genes with highest degree values in each subclass of IDEN-P as the hub genes. We also compared the expression levels of these genes in the disease and normal states using three datasets (GSE29429, GSE69581 and GSE83456). In Common-P,

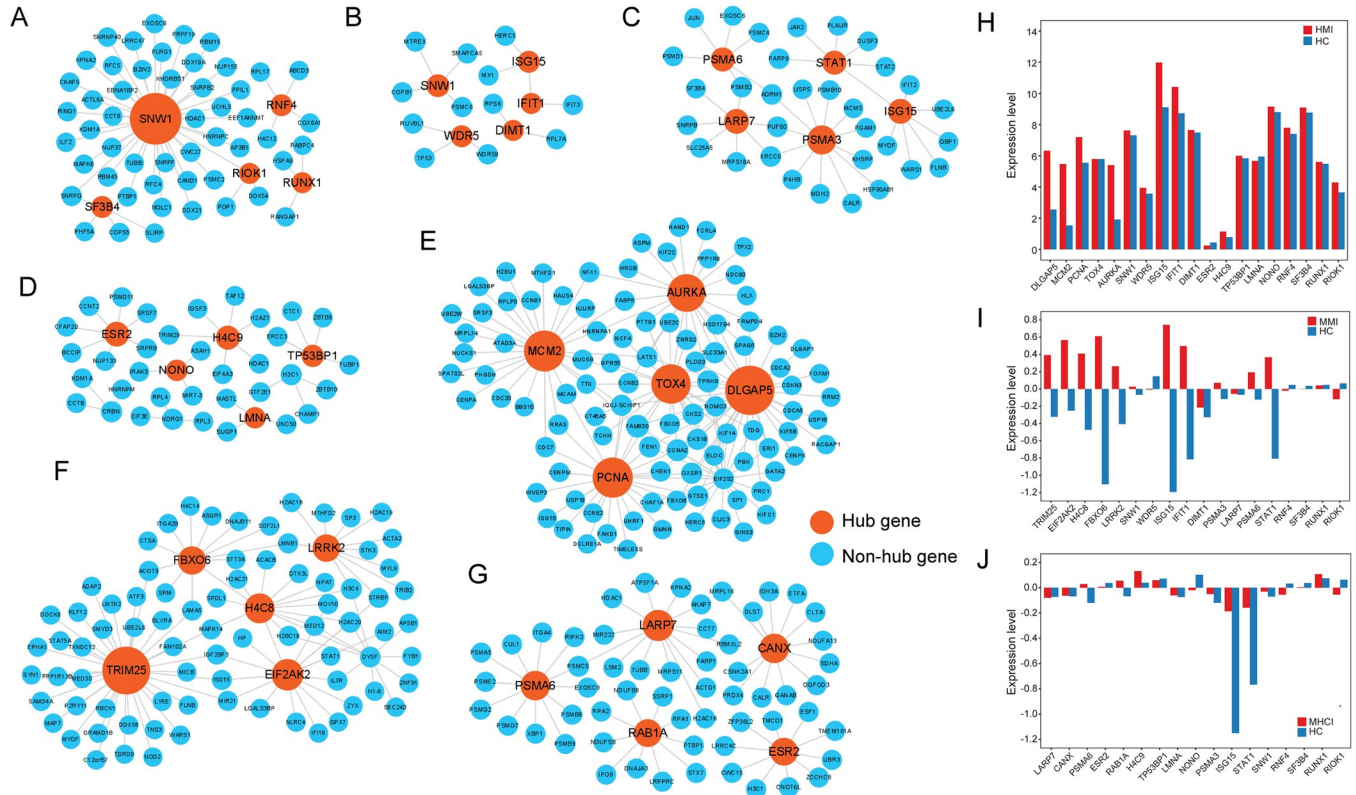


Fig 9. Analysis of hub genes in different subnetworks. (A) Hub genes in Common-P. (B) Hub genes in HMI-MMI-SP. (C) Hub genes in MMI-MHCI-SP. (D) Hub genes in HMI-MHCI-SP. (E) Hub genes in HMI-SP. (F) Hub genes in MMI-SP. (G) Hub genes in MHCI-SP. (H–J) Expression levels of hub genes in the disease and healthy states.

<https://doi.org/10.1371/journal.pcbi.1010744.g009>

ISG15 and *IFIT1* were two hub genes in HMI-MMI-SP (Fig 9B). *ISG15* could inhibit the replication and release of viruses [62]. As an interferon-induced protein, *IFIT1* plays a key role in immune responses [63]. The expression levels of *ISG15* and *IFIT1* were remarkably higher in the disease states, probably because these two interferon-related genes are abundantly expressed for strengthening immune defenses. Two proteasome-related proteins, namely *PSMA3* and *PSMA6*, were hub genes in MMI-MHCI-SP, and their higher expression in patients suggested that the host may utilize proteasomes to eliminate disease-related proteins (Fig 9C). Three hub genes (*TP53BP1*, *LMNA* and *NONO*) in HMI-MHCI-SP are involved in DNA damage repair and thus maintain the genome stability [64–66] (Fig 9D). Existing studies reported that overexpression of estrogen receptor (e.g. *ESR2*) could suppress viral replication [67]. Collectively, the hub genes shared by the two states showed the ability to enhance host immunity (S7 Table).

Most viruses invade hosts by regulating the cell cycle. The hub genes (*DLGAP5*, *MCM2*, *PCNA* and *AURKA*) in HMI-SP consistently encode cell cycle regulatory proteins that ensure proper chromosome segregation [68–72] (Fig 9E). Their higher expression levels in HMI were probably caused by HIV invasion (Fig 9H), resulting in abnormal cell cycle regulation and pathological forms of mitosis. Endoplasmic reticulum stress (ERS) is a critical cellular self-protective mechanism [73]. The hub gene *TRIM25* in MMI-SP can provide negative feedback regulation of ERS [74] (Fig 9F). We speculated that MTB may use *TRIM25* to respond to the ERS to accelerate the progression of TB disease. Moreover, *FBXO6* and *EIF2AK2* are also involved in ERS or cellular homeostatic responses [75,76]. The higher expression of these three genes in

MMI supported the above speculation (Fig 9I). Transcriptional regulation controls the normal expression of genes, and disruption of this process may lead to disease [77]. In MHCI-SP, four hub genes (*LARP7*, *ESR2*, *CANX* and *RAB1A*) were reported to be involved in transcriptional regulation and translational processes [5,78–80] (Fig 9G). In contrast to the hub genes shared by two states, those specific to each state could facilitate pathogen infection (S7 Table).

Drug repurposing based on gene pairs

Here, we evaluated the three interaction-based proximity metrics proposed in this work. Furthermore, we compared our algorithms with existing network-based drug repurposing methods, including the Guney's distance [19,20], Zhou's distance [17], diffusion state distance (DSD) methods [19], and graph kernel-based methods [18]. All competing algorithms explored the relationship between drug targets and disease genes rather than gene pairs. The kernel methods were implemented using in-house scripts, and the other competing approaches were implemented using the codes on GitHub. Among the 2065 approved drugs collected in this study, 38 HIV- and 15 TB-related drugs could be considered as positive samples (S2 Data). If one algorithm assigns the higher rankings for known drugs, this method may provide more effective predictions. The evaluation measures (e.g. AUC, recall and precision at the top rankings) are the same as those used by Santos et al.'s work [18]. As shown in Fig 10 and S8 and S9 Tables, our distance measures performed more favorably than disease gene-based proximity metrics (i.e. Guney's distance and Zhou's distance), suggesting that incorporation of disease-related interactions is indeed useful for drug repurposing. Our methods also showed advantages over other network-based algorithms except the DSD method. Among our three measures, the distance $d_2(P,T)$ achieved generally better performance than the distance $d_1(P,T)$ and the combined distance $d_3(P,T)$. This may be because fixing the bigger gene set (e.g. the PPI set) and computing the shortest distance to the genes in the smaller set (e.g. the drug target set) could more effectively reflect the proximity. Based on the analysis, the distance $d_2(P,T)$ was finally selected for drug repurposing in this work.

By observing the distribution of standardized proximity metrics, we considered that the drugs with a Z-score less than -1.036 may be effective against pathogen infections. We thus obtained 178, 150 and 169 drugs (approximately the top 8% of all drugs) for MMI, HMI and MHCI, respectively (S2 Data). Then, we compared the up- and down-regulated DEGs of each disease state against the drug-induced gene expression profiles in CMAP. The GSEA score provided by CMAP could be used as an auxiliary indication to validate the above candidates (a negative score was preferred). Moreover, we ensured that each candidate has anti-pathogen evidences by manually checking the literature. Finally, we retained drugs having fewer side effects and excluded nutritional drugs, metal drugs (e.g. copper and zinc) and radioligand diagnostic agents. According to the above criteria, 10, 10 and 11 drugs were chosen for MMI, HMI and MHCI, respectively.

Repurposable drug candidates for MMI. Among the 10 identified drugs, several candidates have been applied to the treatment of bacterial diseases (Fig 11A and S10 Table). As a macrolide antibiotic, for instance, azithromycin could prevent various bacterial infections [81] (Fig 11B). This drug blocks bacterial protein synthesis by binding to the 50S ribosomal subunit [82]. Moxifloxacin, a broad-spectrum fluoroquinolone antibiotic, functions by inhibiting topoisomerases II and IV, therefore interfering with bacterial DNA replication, transcription and repair [83] (Fig 11C). Additionally, ciprofloxacin, sparfloxacin and ofloxacin are also fluoroquinolones and may exhibit similar anti-TB effects. All these four drugs could mediate two gene pairs, *TOP2B-DAXX* and *VHL-CUL5* (Fig 11C). Imatinib is a tyrosine kinase inhibitor for the treatment of malignant tumors. This agent inhibits tyrosine kinases *ABL1* and *ABL2*,

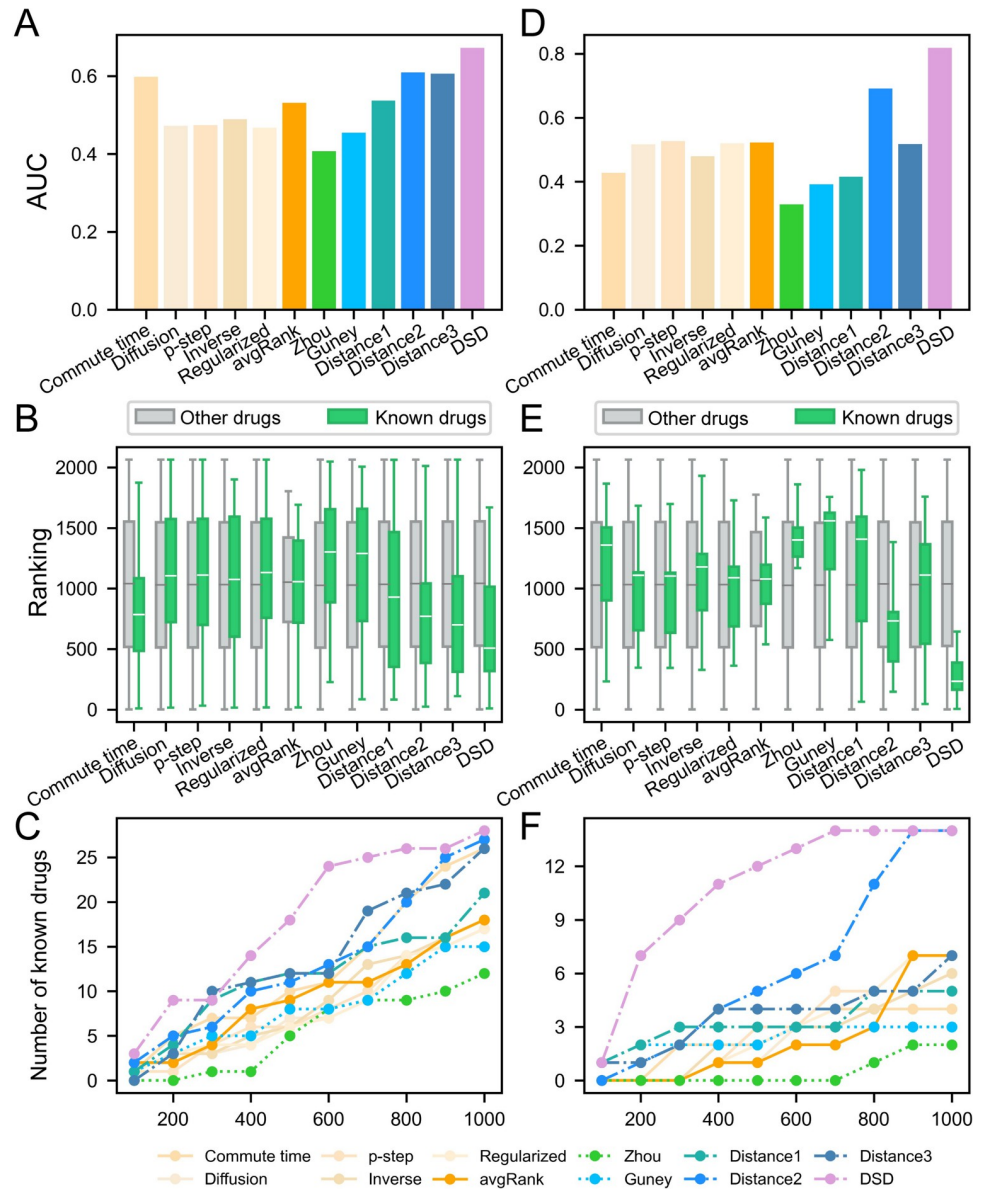


Fig 10. Comparison of our proximity measures and existing network-based drug repurposing methods. (A-C) Performance of different methods on known HIV drugs. (D-F) Performance of different methods on known TB drugs. Evaluation measures include the AUC for identifying known drugs, rankings of known drugs and other drugs, and numbers of known drugs appearing in the top K drugs. These algorithms include our gene pair-based distance measures (distance1, distance2 and distance3), disease gene-based distance measures (Guney’s distance and Zhou’s distance), graph kernel methods (commute time, diffusion, p-step, regularized Laplacian, inverse cosine and avgRank), and diffusion state distance methods (DSD).

<https://doi.org/10.1371/journal.pcbi.1010744.g010>

remarkably reducing the bacterial load in MTB-infected mice [84,85]. Our analysis showed that imatinib may adopt drug targets *KIT* and *PDGFRB* to regulate *JAK2-STAT1*, of which *STAT1* is a key factor in the interferon signaling pathway to resist MTB infection [86] (Fig 11D). Trimethoprim is an anti-folate antibiotic that suppresses bacterial dihydrofolate reductase (*DHFR*), thereby blocking bacterial DNA synthesis and ultimately inhibiting bacterial survival (Fig 11E). Several studies showed that trimethoprim combined with sulfamethoxazole has strong antibacterial activity against MTB in clinical treatment [87–89].

MCM2 can effectively block HIV-1 infection of macrophages by inhibiting viral DNA synthesis [70,93] (S10B Fig). Angiotensin receptor blockers (ARBs) have been reported to combat viral infections (e.g. human coronaviruses) [94,95]. As a representative ARB, irbesartan is mainly used for treating hypertension and diabetic nephropathy. This candidate is highly associated with HIV-related PPIs, such as *NOLC1-SNW1* and *SF3B4-COPS5* [96,97] (S10C Fig). In principle, disrupting the key stages of the HIV life cycle could resist HIV infection. Six drug candidates (i.e. fludarabine, clofarabine, cladribine, gemcitabine, topotecan and irinotecan) could inhibit DNA or RNA production, thereby blocking the transcription or translation process of HIV [98–100]. Estradiol is a sex hormone having anti-inflammatory properties, which could protect nerve cells from injuries and chronic diseases, including neuroAIDS [101,102]. In HIV-1 infected cells, estradiol may inhibit HIV-1 replication by directly altering its transcriptional activation [103,104]. Through targets *ESR1*, *ESR2* and *NR1I2*, estradiol could mediate *NCOA6-NCOA2*, of which *NCOA6* regulates the expression of HIV-1 long terminal repeats, and *NCOA2* serves as a coactivator for HIV-1 Tat proteins [105,106] (S10D Fig).

Repurposable drug candidates for MHCI. The co-infection is not a simple superposition of two mono-infections, and there may be a more complex pathogenesis behind MHCI. Therefore, we should develop rational drugs for the specific pathogenic mechanisms of co-infection in addition to inhibiting the mono-infection of the two pathogens. Among the 11 identified drugs, ciprofloxacin, trimethoprim and toremifene were proposed to resist MTB infection, while topotecan, irinotecan and irbesartan were identified to prevent HIV infection (S11A Fig and S12 Table). Moreover, we obtained another three drugs that were specific to MHCI. Vinorelbine and vinblastine are both anti-mitotic chemotherapeutic agents that block microtubule polymerization to stop cell division in mitosis [107,108] (S11B Fig). These two drugs have been applied to the treatment of HIV-1 patients. The inflammatory response caused by pathogen infections is a primary cause of tissue damage and death [109]. Quercetin, widely distributed in fruits and vegetables, is a flavonoid compound with anti-oxidant and anti-inflammatory effects [110] (S11C Fig). Quercetin could function as the inhibitor of HIV reverse transcriptase and suppress the activities of various cellular DNA and RNA polymerases [111]. In vitro cell assays also showed that quercetin exhibits modest anti-HIV activities [112]. Additionally, extensive studies reported that quercetin has certain efficacy against TB [113–115]. For example, quercetin could inhibit a variety of enzymes (e.g. isocitrate lyase and hyaluronidase) that are essential for the growth and survival of MTB [116]. Our predictions implied that it could also have a host-directed mechanism of action and is a promising drug for the treatment of co-infection.

Discussion

TB and AIDS are the two most lethal infectious diseases which are closely interconnected in the world. Although some efforts have been devoted to exploring the pathogeneses of MMI, HMI and MHCI, the similarities and specificities of the host responses to these infections have not been systematically compared and analyzed. To this end, our study designed a computational framework named IDEN to construct differential expression networks for the three disease states by combining differential and non-differential interactions, which were generated by the DCA and DEA methods, respectively. This framework identified 1007 genes and 911 gene pairs for MMI, 1266 genes and 1314 gene pairs for HMI and 1576 genes and 1363 gene pairs for MHCI. The identified genes were significantly enriched in known host targets of MTB and HIV as well as in functionally important gene sets and could be adopted as signatures to discriminate between disease and normal samples. We then compared the three networks from different viewpoints. At the genomic level, MMI-G and MHCI-G had the highest enrichment ratio on a greater number of chromosomes compared to HMI-G, and a similar

trend was shown for the associated SNPs, dSNPs and disease-causing genes in MMI. Moreover, we found that genes shared by more statuses tended to be more evolutionarily conserved, posttranslationally modified and topologically important. Gene expression analyses showed that Common-G showed the highest expression level and HMI-SG held the lowest expression level. Regarding the gene pairs, HMI-SP yielded higher expression correlations than the other specific groups, while the overlapping pairs associated with co-infections had significantly lower correlations in the disease states. The patterns of expression correlation changes of Common-P revealed that MHCI may be more similar to MMI. Gene enrichment analyses exhibited that MMI-G and MHCI-G shared more pathways and biological processes, further illustrating that the pathogenic mechanism of TB might play a dominant role in the progression of MHCI. Functional analyses of the hub nodes in subnetworks suggested that the hub genes specific to each disease state may promote pathogen infections, while the hub genes shared by two disease states may enhance immune responses. Regarding drug repurposing, the network-based strategy has been successfully applied to this field, but existing works only considered the relationship between known drug targets and disease-related proteins rather than disease-related PPIs. We thus improved the existing proximity measure by focusing on gene pairs detected in this work. Finally, approximately ten reusable drugs were identified for each disease state, which may provide new clues to the therapy for relevant diseases.

Despite the results derived from this study, there is still room for improvement in the future work. First, the available datasets of relevant blood expression profiles were relatively limited. Although we designed the customized IDEN method for the current datasets, this framework could be improved by integrating other transcriptomic data (e.g. RNA-seq). Additional datasets could further validate the reliability of current results. Second, we only used the interactions from HIPPIE to build the reference PPI network of IDEN, which may lead to false positives in the identified gene pairs. To solve this problem, we could use a more rigorous threshold for confidence scores in HIPPIE or combine other PPI resources (e.g. PDB database) to strengthen the quality of human PPI network. Third, we used consistent PCC thresholds in the second and third filtering steps to detect DIs for different disease states (S2 Fig), which may lose gene pairs specific to a certain state. In future we could design adaptive thresholds for each state to retrieve DIs. Fourth, the drug-target network for drug repurposing only included records from DrugBank and TTD. Integration of more drug-related resources (e.g. PharmGKB, ChEMBL and BindingDB databases) may improve the coverage of drug-target interactions, thereby providing more insights into reusable drugs. Fifth, although most proposed drugs in this work have been supported by comprehensive evidences, all drugs must undergo rigorous randomized clinical trials before they can be used to treat patients. Sixth, in addition to AIDS, other diseases (e.g. diabetes and hypertension) were interconnected with TB. We could conduct the similar comparison and analysis for other complex disease states. In conclusion, we provide an integrative framework for the identification of important gene pairs and reusable drugs for the infections caused by MTB and/or HIV, which may help deepen our understanding of the mechanisms underlying the interactions between the host and pathogens.

Supporting information

S1 Fig. Normalized expression profiles of randomly selected samples in each dataset. (PDF)

S2 Fig. Process of extracting DIs by the DCA method. (A) Extraction of DIs for HMI. (B) Extraction of DIs for MMI. (C) Extraction of DIs for MHCI. (PDF)

S3 Fig. Samples in GSE69581 and GSE83456 datasets before and after removal of batch effects. (A) Normalized expression profiles of samples. (B) Principal component analysis of samples. Left: results before batch effect removal, middle: results after batch effect removal and right: results after quantile normalization.

(PDF)

S4 Fig. Volcano plot of DEGs for three disease states. (A) DEGs of HMI. (B) DEGs of MMI. (C) DEGs of MHCI.

(PDF)

S5 Fig. Classification of genes and gene pairs identified by IDEN framework. (A) Genes identified by the IDEN (IDEN-G). (B) Gene pairs identified by the IDEN (IDEN-P). (C) Seven subclasses of IDEN-G or IDEN-P.

(PDF)

S6 Fig. Genomic features of disease-related genes. (A) Chromosomal distribution of detected genes. (B) Chromosomal distribution of SNP density of detected genes. (C) Proportion of disease-related SNPs among total SNPs of detected genes. (D) Proportion of disease-causing genes among detected genes.

(PDF)

S7 Fig. Chromosomal distance of disease-related genes in different subclasses.

(PDF)

S8 Fig. Expression level of disease-related genes in different groups.

(PDF)

S9 Fig. Gene ontology analysis of disease-related genes. (A) MMI-G. (B) HMI-G. (C) MHCI-G. (D) Common-G.

(PDF)

S10 Fig. Drug-target-PPI associations for HMI. (A) Relationship between anti-HMI drug candidates and HMI-associated PPIs. Top 10 PPIs with the highest number of shortest links to drug targets are shown (one PPI may correspond to multiple targets). (B) Closest PPIs regulated by raloxifene. (C) Closest PPIs regulated by irbesartan. (D) Closest PPIs regulated by estradiol.

(PDF)

S11 Fig. Drug-target-PPI associations for MHCI. (A) Relationship between anti-MHCI drug candidates and MHCI-associated PPIs. Top 10 PPIs with the highest number of shortest links to drug targets are shown (one PPI may correspond to multiple targets). (B) Closest PPIs regulated by vinblastine and vinorelbine. (C) Closest PPIs regulated by quercetin.

(PDF)

S1 Table. Gene expression profile datasets used in this study.

(PDF)

S2 Table. Overlap between identified genes and existing resources.

(PDF)

S3 Table. Number of IDEN genes enriched in functionally important gene sets.

(PDF)

S4 Table. Number of subclass genes enriched in functionally important gene sets.

(PDF)

S5 Table. Performance of genes and gene pairs for classification of different samples.
(PDF)

S6 Table. AUC values achieved by different gene signatures.
(PDF)

S7 Table. Biological functions of hub genes in different subnetworks.
(PDF)

S8 Table. Recall and precision at the top K drugs of different methods for HIV drugs.
(PDF)

S9 Table. Recall and precision at the top K drugs of different methods for TB drugs.
(PDF)

S10 Table. Repurposed anti-MMI drug candidates.
(PDF)

S11 Table. Repurposed anti-HMI drug candidates.
(PDF)

S12 Table. Repurposed anti-MHCI drug candidates.
(PDF)

S1 Data. Original data used for generating figures.
(XLSX)

S2 Data. Original data used for drug repurposing.
(XLSX)

Author Contributions

Conceptualization: Rong Liu.

Data curation: Yao Jiang.

Formal analysis: Yao Jiang, Jia-Xuan Zhang, Rong Liu.

Funding acquisition: Rong Liu.

Investigation: Yao Jiang, Jia-Xuan Zhang, Rong Liu.

Methodology: Yao Jiang, Rong Liu.

Project administration: Rong Liu.

Resources: Rong Liu.

Software: Yao Jiang.

Supervision: Rong Liu.

Validation: Yao Jiang, Jia-Xuan Zhang, Rong Liu.

Visualization: Yao Jiang, Jia-Xuan Zhang, Rong Liu.

Writing – original draft: Yao Jiang, Rong Liu.

Writing – review & editing: Rong Liu.

References

1. Yang L, Hu X, Chai X, Ye Q, Pang J, Li D, et al. Opportunities for overcoming tuberculosis: Emerging targets and their inhibitors. *Drug Discov Today*. 2022; 27(1):326–36. <https://doi.org/10.1016/j.drudis.2021.09.003> PMID: 34537334
2. Bell LCK, Noursadeghi M. Pathogenesis of HIV-1 and Mycobacterium tuberculosis co-infection. *Nat Rev Microbiol*. 2018; 16(2):80–90. <https://doi.org/10.1038/nrmicro.2017.128> PMID: 29109555
3. McShane H. Co-infection with HIV and TB: double trouble. *Int J STD AIDS*. 2005; 16(2):95–100; quiz 1. <https://doi.org/10.1258/0956462053057576> PMID: 15807935
4. Toossi Z. Virological and immunological impact of tuberculosis on human immunodeficiency virus type 1 disease. *J Infect Dis*. 2003; 188(8):1146–55. <https://doi.org/10.1086/378676> PMID: 14551885
5. A. K, A.J. G-P, AM. M. HIV-associated pediatric tuberculosis: prevention, diagnosis and treatment. *Curr Opin HIV AIDS*. 2018; 13(6):501–6. <https://doi.org/10.1097/COH.0000000000000500> PMID: 30286040
6. Berry MP, Graham CM, McNab FW, Xu Z, Bloch SA, Oni T, et al. An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature*. 2010; 466(7309):973–7. <https://doi.org/10.1038/nature09247> PMID: 20725040
7. Kaforou M, Wright VJ, Oni T, French N, Anderson ST, Bangani N, et al. Detection of tuberculosis in HIV-infected and -uninfected African adults using whole blood RNA expression signatures: a case-control study. *PLoS Med*. 2013; 10(10):e1001538. <https://doi.org/10.1371/journal.pmed.1001538> PMID: 24167453
8. Devadas K, Biswas S, Haleyrigirisetty M, Wood O, Ragupathy V, Lee S, et al. Analysis of Host Gene Expression Profile in HIV-1 and HIV-2 Infected T-Cells. *PLoS One*. 2016; 11(1):e0147421. <https://doi.org/10.1371/journal.pone.0147421> PMID: 26821323
9. Chen Y, Wang Q, Lin S, Lai J, Lin J, Ao W, et al. Meta-Analysis of Peripheral Blood Transcriptome Datasets Reveals a Biomarker Panel for Tuberculosis in Patients Infected With HIV. *Front Cell Infect Microbiol*. 2021; 11:585919. <https://doi.org/10.3389/fcimb.2021.585919> PMID: 33816327
10. Duffy FJ, Thompson EG, Scriba TJ, Zak DE. Multinomial modelling of TB/HIV co-infection yields a robust predictive signature and generates hypotheses about the HIV+TB+ disease state. *PLoS One*. 2019; 14(7):e0219322. <https://doi.org/10.1371/journal.pone.0219322> PMID: 31306460
11. Bandyopadhyay S, Mehta M, Kuo D, Sung MK, Chuang R, Jaehnig EJ, et al. Rewiring of genetic networks in response to DNA damage. *Science*. 2010; 330(6009):1385–9. <https://doi.org/10.1126/science.1195618> PMID: 21127252
12. Sambarey A, Devaprasad A, Mohan A, Ahmed A, Nayak S. Unbiased Identification of Blood-based Biomarkers for Pulmonary Tuberculosis by Modeling and Mining Molecular Interaction Networks. *Ebiomedicine*. 2017. <https://doi.org/10.1016/j.ebiom.2016.12.009> PMID: 28065665
13. Kumar D, Nath L, Kamal MA, Varshney A, Jain A, Singh S, et al. Genome-wide analysis of the host intracellular network that regulates survival of Mycobacterium tuberculosis. *Cell*. 2010; 140(5):731–43. <https://doi.org/10.1016/j.cell.2010.02.012> PMID: 20211141
14. Yoon D, Kim H, Suh-Kim H, Park RW, Lee K. Differentially co-expressed interacting protein pairs discriminate samples under distinct stages of HIV type 1 infection. *BMC Syst Biol*. 2011; 5 Suppl 2(Suppl 2):S1. <https://doi.org/10.1186/1752-0509-5-S2-S1> PMID: 22784566
15. Sun SY, Liu ZP, Zeng T, Wang Y, Chen L. Spatio-temporal analysis of type 2 diabetes mellitus based on differential expression networks. *Sci Rep*. 2013; 3:2268. <https://doi.org/10.1038/srep02268> PMID: 23881262
16. Sun J, Shi Q, Chen X, Liu R. Decoding the similarities and specific differences between latent and active tuberculosis infections based on consistently differential expression networks. *Brief Bioinform*. 2020; 21(6):2084–98. <https://doi.org/10.1093/bib/bbz127> PMID: 31724702
17. Zhou Y, Hou Y, Shen J, Mehra R, Kallianpur A, Culver DA, et al. A network medicine approach to investigation and population-based validation of disease manifestations and drug repurposing for COVID-19. *PLoS Biol*. 2020; 18(11):e3000970. <https://doi.org/10.1371/journal.pbio.3000970> PMID: 33156843
18. Santos SS, Torres M, Galeano D, Sánchez MDM, Cernuzzi L, Paccanaro A. Machine learning and network medicine approaches for drug repositioning for COVID-19. *Patterns (N Y)*. 2022; 3(1):100396. <https://doi.org/10.1016/j.patter.2021.100396> PMID: 34778851
19. Morselli Gysi D, de Valle Í, Zitnik M, Ameli A, Gan X, Varol O, et al. Network medicine framework for identifying drug-repurposing opportunities for COVID-19. *Proceedings of the National Academy of Sciences*. 2021; 118(19):e2025581118. <https://doi.org/10.1073/pnas.2025581118> PMID: 33906951

20. Cheng F, Desai RJ, Handy DE, Wang R, Schneeweiss S, Barabási AL, et al. Network-based approach to prediction and population-based validation of in silico drug repurposing. *Nat Commun.* 2018; 9(1):2691. <https://doi.org/10.1038/s41467-018-05116-5> PMID: 30002366
21. Guney E, Menche J, Vidal M, Barabasi AL. Network-based in silico drug efficacy screening. *Nat Commun.* 2016; 7:10331. <https://doi.org/10.1038/ncomms10331> PMID: 26831545
22. Zhou Y, Hou Y, Shen J, Huang Y, Martin W, Cheng F. Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell Discov.* 2020; 6:14. <https://doi.org/10.1038/s41421-020-0153-3> PMID: 32194980
23. Peng Y, Yuan M, Xin J, Liu X, Wang J. Screening novel drug candidates for Alzheimer's disease by an integrated network and transcriptome analysis. *Bioinformatics.* 2020; 36(17):4626–32. <https://doi.org/10.1093/bioinformatics/btaa563> PMID: 32516365
24. Pickett EK, Herrmann AG, McQueen J, Abt K, Dando O, Tulloch J, et al. Amyloid Beta and Tau Cooperate to Cause Reversible Behavioral and Transcriptional Deficits in a Model of Alzheimer's Disease. *Cell Rep.* 2019; 29(11):3592–604.e5. <https://doi.org/10.1016/j.celrep.2019.11.044> PMID: 31825838
25. Xiao T, Cai Y, Chen B. HIV-1 Entry and Membrane Fusion Inhibitors. *Viruses.* 2021; 13(5). <https://doi.org/10.3390/v13050735> PMID: 33922579
26. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 2002; 30(1):207–10. <https://doi.org/10.1093/nar/30.1.207> PMID: 11752295
27. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015; 43(7):e47. <https://doi.org/10.1093/nar/gkv007> PMID: 25605792
28. Alanis-Lobato G, Andrade-Navarro MA, Schaefer MH. HIPPIE v2.0: enhancing meaningfulness and reliability of protein-protein interaction networks. *Nucleic Acids Res.* 2017; 45(D1):D408–d14. <https://doi.org/10.1093/nar/gkw985> PMID: 27794551
29. Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, Tzur D, et al. DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.* 2008; 36(Database issue):D901–6. <https://doi.org/10.1093/nar/gkm958> PMID: 18048412
30. Zhu F, Han B, Kumar P, Liu X, Ma X, Wei X, et al. Update of TTD: Therapeutic Target Database. *Nucleic Acids Res.* 2010; 38(Database issue):D787–91. <https://doi.org/10.1093/nar/gkp1014> PMID: 19933260
31. Hart T, Chandrashekhar M, Aregger M, Steinhart Z, Brown KR, MacLeod G, et al. High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. *Cell.* 2015; 163(6):1515–26. <https://doi.org/10.1016/j.cell.2015.11.015> PMID: 26627737
32. Eisenberg E, Levanon EY. Human housekeeping genes, revisited. *Trends Genet.* 2013; 29(10):569–74. <https://doi.org/10.1016/j.tig.2013.05.010> PMID: 23810203
33. Wattam AR, Abraham D, Dalay O, Disz TL, Driscoll T, Gabbard JL, et al. PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res.* 2014; 42(Database issue):D581–91. <https://doi.org/10.1093/nar/gkt1099> PMID: 24225323
34. Ammari MG, Gresham CR, McCarthy FM, Nanduri B. HPIDB 2.0: a curated database for host-pathogen interactions. *Database (Oxford).* 2016; 2016. <https://doi.org/10.1093/database/baw103> PMID: 27374121
35. Guirimand T, Delmotte S, Navratil V. VirHostNet 2.0: surfing on the web of virus/host molecular interactions data. *Nucleic Acids Res.* 2015; 43(Database issue):D583–7. <https://doi.org/10.1093/nar/gku1121> PMID: 25392406
36. Durmuş Tekir S, Çakır T, Ardiç E, Sayılırbaş AS, Konuk G, Konuk M, et al. PHISTO: pathogen-host interaction search tool. *Bioinformatics.* 2013; 29(10):1357–8. <https://doi.org/10.1093/bioinformatics/btt137> PMID: 23515528
37. Cosmin L, Stijn M, Jonatan T, David S, Alain C, Colin M, et al. Batch effect removal methods for microarray gene expression data integration: a survey. *Briefings in Bioinformatics.* 2012;(4):4.
38. Kinsella RJ, Kähäri A, Haider S, Zamora J, Proctor G, Spudich G, et al. Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database (Oxford).* 2011; 2011:bar030. <https://doi.org/10.1093/database/bar030> PMID: 21785142
39. Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, et al. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* 2016; 44(D1):D862–8. <https://doi.org/10.1093/nar/gkv1222> PMID: 26582918
40. Piñero J, Queralt-Rosinach N, Bravo À, Deu-Pons J, Bauer-Mehren A, Baron M, et al. DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. *Database (Oxford).* 2015; 2015:bav028. <https://doi.org/10.1093/database/bav028> PMID: 25877637

41. Capra JA, Williams AG, Pollard KS. ProteinHistorian: tools for the comparative analysis of eukaryote protein origin. *PLoS Comput Biol.* 2012; 8(6):e1002567. <https://doi.org/10.1371/journal.pcbi.1002567> PMID: 22761559
42. Sayers EW, Beck J, Bolton EE, Bourexis D, Brister JR, Canese K, et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 2021; 49(D1):D10–d7. <https://doi.org/10.1093/nar/gkaa892> PMID: 33095870
43. Hornbeck PV, Zhang B, Murray B, Kornhauser JM, Latham V, Skrzypek E. PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res.* 2015; 43(Database issue):D512–20. <https://doi.org/10.1093/nar/gku1267> PMID: 25514926
44. Enright AJ, Van Dongen S, Ouzounis CA. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* 2002; 30(7):1575–84. <https://doi.org/10.1093/nar/30.7.1575> PMID: 11917018
45. Misselbeck K, Parolo S, Lorenzini F, Savoca V, Leonardelli L, Bora P, et al. A network-based approach to identify deregulated pathways and drug effects in metabolic syndrome. *Nat Commun.* 2019; 10(1):5215. <https://doi.org/10.1038/s41467-019-13208-z> PMID: 31740673
46. Kong J, Lee H, Kim D, Han SK, Ha D, Shin K, et al. Network-based machine learning in colorectal and bladder organoid models predicts anti-cancer drug efficacy in patients. *Nat Commun.* 2020; 11(1):5485. <https://doi.org/10.1038/s41467-020-19313-8> PMID: 33127883
47. Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, et al. A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell.* 2017; 171(6):1437–52.e17. <https://doi.org/10.1016/j.cell.2017.10.049> PMID: 29195078
48. Zhang W, Zeng T, Chen L. EdgeMarker: Identifying differentially correlated molecule pairs as edge-biomarkers. *J Theor Biol.* 2014; 362:35–43. <https://doi.org/10.1016/j.jtbi.2014.05.041> PMID: 24931676
49. Halehalli RR, Nagarajaram HA. Molecular principles of human virus protein-protein interactions. *Bioinformatics.* 2015; 31(7):1025–33. <https://doi.org/10.1093/bioinformatics/btu763> PMID: 25417202
50. Zhang L, Yao J, Withers J, Xin XF, Banerjee R, Fariduddin Q, et al. Host target modification as a strategy to counter pathogen hijacking of the jasmonate hormone receptor. *Proc Natl Acad Sci U S A.* 2015; 112(46):14354–9. <https://doi.org/10.1073/pnas.1510745112> PMID: 26578782
51. Höflmayer D, Willich C, Hube-Magg C, Simon R, Lang D, Neubauer E, et al. SNW1 is a prognostic biomarker in prostate cancer. *Diagn Pathol.* 2019; 14(1):33. <https://doi.org/10.1186/s13000-019-0810-8> PMID: 31043167
52. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OmicS.* 2012; 16(5):284–7. <https://doi.org/10.1089/omi.2011.0118> PMID: 22455463
53. Fu C, Yang S, Yang X, Lian X, Huang Y, Dong X, et al. Human Gene Functional Network-Informed Prediction of HIV-1 Host Dependency Factors. *mSystems.* 2020; 5(6). <https://doi.org/10.1128/mSystems.00960-20> PMID: 33144314
54. Taromi S, Lewens F, Arsenic R, Sedding D, Sängner J, Kunze A, et al. Proteasome inhibitor bortezomib enhances the effect of standard chemotherapy in small cell lung cancer. *Oncotarget.* 2017; 8(57):97061–78. <https://doi.org/10.18632/oncotarget.21221> PMID: 29228593
55. Maddocks S, Scandurra GM, Nourse C, Bye C, Williams RB, Slobedman B, et al. Gene expression in HIV-1/Mycobacterium tuberculosis co-infected macrophages is dominated by M. tuberculosis. *Tuberculosis (Edinb).* 2009; 89(4):285–93. <https://doi.org/10.1016/j.tube.2009.05.003> PMID: 19520608
56. Fallahi-Sichani M, Kirschner DE, Linderman JJ. NF- κ B Signaling Dynamics Play a Key Role in Infection Control in Tuberculosis. *Front Physiol.* 2012; 3:170.
57. Khan KA, Abbas W, Varin A, Kumar A, Di Martino V, Dichamp I, et al. HIV-1 Nef interacts with HCV Core, recruits TRAF2, TRAF5 and TRAF6, and stimulates HIV-1 replication in macrophages. *J Innate Immun.* 2013; 5(6):639–56. <https://doi.org/10.1159/000350517> PMID: 23774506
58. Häkli M, Lorick KL, Weissman AM, Jänne OA, Palvimo JJ. Transcriptional coregulator SNURF (RNF4) possesses ubiquitin E3 ligase activity. *FEBS Lett.* 2004; 560(1–3):56–62. [https://doi.org/10.1016/S0014-5793\(04\)00070-5](https://doi.org/10.1016/S0014-5793(04)00070-5) PMID: 14987998
59. Galanty Y, Belotserkovskaya R, Coates J, Jackson SP. RNF4, a SUMO-targeted ubiquitin E3 ligase, promotes DNA double-strand break repair. *Genes Dev.* 2012; 26(11):1179–95. <https://doi.org/10.1101/gad.188284.112> PMID: 22661229
60. Golas MM, Sander B, Will CL, Lührmann R, Stark H. Molecular architecture of the multiprotein splicing factor SF3b. *Science.* 2003; 300(5621):980–4. <https://doi.org/10.1126/science.1084155> PMID: 12738865
61. Shen Q, Eun JW, Lee K, Kim HS, Yang HD, Kim SY, et al. Barrier to autointegration factor 1, procollagen-lysine, 2-oxoglutarate 5-dioxygenase 3, and splicing factor 3b subunit 4 as early-stage cancer

- decision markers and drivers of hepatocellular carcinoma. *Hepatology*. 2018; 67(4):1360–77. <https://doi.org/10.1002/hep.29606> PMID: 29059470
62. Perng YC, Lenschow DJ. ISG15 in antiviral immunity and beyond. *Nat Rev Microbiol*. 2018; 16(7):423–39. <https://doi.org/10.1038/s41579-018-0020-5> PMID: 29769653
 63. Diamond MS. IFIT1: A dual sensor and effector molecule that detects non-2'-O methylated viral RNA and inhibits its translation. *Cytokine Growth Factor Rev*. 2014; 25(5):543–50. <https://doi.org/10.1016/j.cytogfr.2014.05.002> PMID: 24909568
 64. Karhumaa P, Parkkila S, Waheed A, Parkkila AK, Kaunisto K, Tucker PW, et al. Nuclear NonO/p54 (nrb) protein is a nonclassical carbonic anhydrase. *J Biol Chem*. 2000; 275(21):16044–9. <https://doi.org/10.1074/jbc.275.21.16044> PMID: 10821857
 65. Davidson PM, Lammerding J. Broken nuclei—lamins, nuclear mechanics, and disease. *Trends Cell Biol*. 2014; 24(4):247–56. <https://doi.org/10.1016/j.tcb.2013.11.004> PMID: 24309562
 66. Furuie H, Arimura-Omori M, Hamada N, Yanagihara T, Kiyohara C. The Association of Aging-Related Polymorphisms with Susceptibility to Lung Cancer: A Case-Control Study in a Japanese Population. *Asian Pac J Cancer Prev*. 2021; 22(4):1279–85. <https://doi.org/10.31557/APJCP.2021.22.4.1279> PMID: 33906323
 67. Lasso G, Mayer SV, Winkelmann ER, Chu T, Elliot O, Patino-Galindo JA, et al. A Structure-Informed Atlas of Human-Virus Interactions. *Cell*. 2019; 178(6):1526–41.e16. <https://doi.org/10.1016/j.cell.2019.08.005> PMID: 31474372
 68. Herce HD, Rajan M, Lättig-Tünnemann G, Fillies M, Cardoso MC. A novel cell permeable DNA replication and repair marker. *Nucleus*. 2014; 5(6):590–600. <https://doi.org/10.4161/nucl.36290> PMID: 25484186
 69. Damodaran AP, Vaufrey L, Gavard O, Prigent C. Aurora A Kinase Is a Priority Pharmaceutical Target for the Treatment of Cancers. *Trends Pharmacol Sci*. 2017; 38(8):687–700. <https://doi.org/10.1016/j.tips.2017.05.003> PMID: 28601256
 70. Mlcochova P, Sutherland KA, Watters SA, Bertoli C, de Bruin RA, Rehwinkel J, et al. A G1-like state allows HIV-1 to bypass SAMHD1 restriction in macrophages. *EMBO J*. 2017; 36(5):604–16. <https://doi.org/10.15252/embj.201696025> PMID: 28122869
 71. Bertolin G, Tramier M. Insights into the non-mitotic functions of Aurora kinase A: more than just cell division. *Cell Mol Life Sci*. 2020; 77(6):1031–47. <https://doi.org/10.1007/s00018-019-03310-2> PMID: 31562563
 72. Zheng R, Shi Z, Li W, Yu J, Wang Y, Zhou Q. Identification and prognostic value of DLGAP5 in endometrial cancer. *PeerJ*. 2020; 8:e10433. <https://doi.org/10.7717/peerj.10433> PMID: 33312770
 73. Dicks N, Gutierrez K, Michalak M, Bordignon V, Agellon LB. Endoplasmic reticulum stress, genome damage, and cancer. *Front Oncol*. 2015; 5:11. <https://doi.org/10.3389/fonc.2015.00011> PMID: 25692096
 74. Liu Y, Tao S, Liao L, Li Y, Li H, Li Z, et al. TRIM25 promotes the cell survival and growth of hepatocellular carcinoma through targeting Keap1-Nrf2 pathway. *Nat Commun*. 2020; 11(1):348. <https://doi.org/10.1038/s41467-019-14190-2> PMID: 31953436
 75. Chen X, Duan LH, Luo PC, Hu G, Yu X, Liu J, et al. FBXO6-Mediated Ubiquitination and Degradation of Ero1L Inhibits Endoplasmic Reticulum Stress-Induced Apoptosis. *Cell Physiol Biochem*. 2016; 39(6):2501–8. <https://doi.org/10.1159/000452517> PMID: 27855403
 76. Cesaro T, Hayashi Y, Borghese F, Vertommen D, Wavreil F, Michiels T. PKR activity modulation by phosphomimetic mutations of serine residues located three aminoacids upstream of double-stranded RNA binding motifs. *Sci Rep*. 2021; 11(1):9188. <https://doi.org/10.1038/s41598-021-88610-z> PMID: 33911136
 77. Lee TI, Young RA. Transcriptional regulation and its misregulation in disease. *Cell*. 2013; 152(6):1237–51. <https://doi.org/10.1016/j.cell.2013.02.014> PMID: 23498934
 78. Karar J, Dolt KS, Qadar Pasha MA. Endoplasmic reticulum stress response in murine kidney exposed to acute hypobaric hypoxia. *FEBS Lett*. 2008; 582(17):2521–6. <https://doi.org/10.1016/j.febslet.2008.06.019> PMID: 18572019
 79. Markert A, Grimm M, Martinez J, Wiesner J, Meyerhans A, Meyuhans O, et al. The La-related protein LARP7 is a component of the 7SK ribonucleoprotein and affects transcription of cellular and viral polymerase II genes. *EMBO Rep*. 2008; 9(6):569–75. <https://doi.org/10.1038/embor.2008.72> PMID: 18483487
 80. Thomas JD, Zhang YJ, Wei YH, Cho JH, Morris LE, Wang HY, et al. Rab1A is an mTORC1 activator and a colorectal oncogene. *Cancer Cell*. 2014; 26(5):754–69. <https://doi.org/10.1016/j.ccell.2014.09.008> PMID: 25446900

81. Zuckerman JM, Qamar F, Bono BR. Macrolides, ketolides, and glycolcyclines: azithromycin, clarithromycin, telithromycin, tigecycline. *Infect Dis Clin North Am*. 2009; 23(4):997–1026, ix-x. <https://doi.org/10.1016/j.idc.2009.06.013> PMID: 19909895
82. Champney WS, Miller M. Inhibition of 50S ribosomal subunit assembly in *Haemophilus influenzae* cells by azithromycin and erythromycin. *Curr Microbiol*. 2002; 44(6):418–24. <https://doi.org/10.1007/s00284-001-0016-6> PMID: 12000992
83. Majalekar PP, Shirote PJ. Fluoroquinolones: Blessings Or Curses. *Curr Drug Targets*. 2020; 21(13):1354–70. <https://doi.org/10.2174/1389450121666200621193355> PMID: 32564750
84. Napier RJ, Rafi W, Cheruvu M, Powell KR, Zaunbrecher MA, Bornmann W, et al. Imatinib-sensitive tyrosine kinases regulate mycobacterial pathogenesis and represent therapeutic targets against tuberculosis. *Cell Host Microbe*. 2011; 10(5):475–85. <https://doi.org/10.1016/j.chom.2011.09.010> PMID: 22100163
85. Bruns H, Stegelmann F, Fabri M, Döhner K, van Zandbergen G, Wagner M, et al. Abelson tyrosine kinase controls phagosomal acidification required for killing of *Mycobacterium tuberculosis* in human macrophages. *J Immunol*. 2012; 189(8):4069–78. <https://doi.org/10.4049/jimmunol.1201538> PMID: 22988030
86. Yi XH, Zhang B, Fu YR, Yi ZJ. STAT1 and its related molecules as potential biomarkers in *Mycobacterium tuberculosis* infection. *J Cell Mol Med*. 2020; 24(5):2866–78. <https://doi.org/10.1111/jcmm.14856> PMID: 32048448
87. Vilchère C, Jacobs WR Jr., The combination of sulfamethoxazole, trimethoprim, and isoniazid or rifampin is bactericidal and prevents the emergence of drug resistance in *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother*. 2012; 56(10):5142–8.
88. Davies Forsman L, Schön T, Simonsson US, Bruchfeld J, Larsson M, Juréen P, et al. Intra- and extracellular activities of trimethoprim-sulfamethoxazole against susceptible and multidrug-resistant *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother*. 2014; 58(12):7557–9. <https://doi.org/10.1128/AAC.02995-14> PMID: 25246405
89. Liu T, Wang B, Guo J, Zhou Y, Julius M, Njire M, et al. Role of folP1 and folP2 Genes in the Action of Sulfamethoxazole and Trimethoprim Against *Mycobacteria*. *J Microbiol Biotechnol*. 2015; 25(9):1559–67. <https://doi.org/10.4014/jmb.1503.03053> PMID: 25907064
90. Yoon YS, Jang Y, Hoenen T, Shin H, Lee Y, Kim M. Antiviral activity of sertindole, raloxifene and ibutamoren against transcription and replication-competent Ebola virus-like particles. *BMB Rep*. 2020; 53(3):166–71. <https://doi.org/10.5483/BMBRep.2020.53.3.175> PMID: 31964466
91. Chiou WC, Hsu MS, Chen YT, Yang JM, Tsay YG, Huang HC, et al. Repurposing existing drugs: identification of SARS-CoV-2 3C-like protease inhibitors. *J Enzyme Inhib Med Chem*. 2021; 36(1):147–53. <https://doi.org/10.1080/14756366.2020.1850710> PMID: 33430659
92. Takeda M, Ikeda M, Mori K, Yano M, Ariumi Y, Dansako H, et al. Raloxifene inhibits hepatitis C virus infection and replication. *FEBS Open Bio*. 2012; 2:279–83. <https://doi.org/10.1016/j.fob.2012.08.003> PMID: 23650611
93. Mlcochova P, Caswell SJ, Taylor IA, Towers GJ, Gupta RK. DNA damage induced by topoisomerase inhibitors activates SAMHD1 and blocks HIV-1 infection of macrophages. *EMBO J*. 2018; 37(1):50–62. <https://doi.org/10.15252/emboj.201796880> PMID: 29084722
94. Erlandson KM, Kitch D, Wester CW, Kalayjian RC, Overton ET, Castillo-Mancilla J, et al. The Impact of Statin and Angiotensin-Converting Enzyme Inhibitor/Angiotensin Receptor Blocker Therapy on Cognitive Function in Adults With Human Immunodeficiency Virus Infection. *Clin Infect Dis*. 2017; 65(12):2042–9. <https://doi.org/10.1093/cid/cix645> PMID: 29020174
95. Seko Y. Effect of the angiotensin II receptor blocker olmesartan on the development of murine acute myocarditis caused by coxsackievirus B3. *Clin Sci (Lond)*. 2006; 110(3):379–86. <https://doi.org/10.1042/CS20050299> PMID: 16336207
96. Brès V, Yoshida T, Pickle L, Jones KA. SKIP interacts with c-Myc and Menin to promote HIV-1 Tat transactivation. *Mol Cell*. 2009; 36(1):75–87. <https://doi.org/10.1016/j.molcel.2009.08.015> PMID: 19818711
97. Kinoshita SM, Krutzik PO, Nolan GP. COP9 signalosome component JAB1/CSN5 is necessary for T cell signaling through LFA-1 and HIV-1 replication. *PLoS One*. 2012; 7(7):e41725. <https://doi.org/10.1371/journal.pone.0041725> PMID: 22911848
98. Montillo M, Ricci F, Tedeschi A. Role of fludarabine in hematological malignancies. *Expert Rev Anticancer Ther*. 2006; 6(9):1141–61. <https://doi.org/10.1586/14737140.6.9.1141> PMID: 17020450
99. Ghanem H, Jabbour E, Faderl S, Ghandhi V, Plunkett W, Kantarjian H. Clofarabine in leukemia. *Expert Rev Hematol*. 2010; 3(1):15–22. <https://doi.org/10.1586/ehm.09.70> PMID: 21082931

100. Schmidt F, Rieger J, Wischhusen J, Naumann U, Weller M. Glioma cell sensitivity to topotecan: the role of p53 and topotecan-induced DNA damage. *Eur J Pharmacol.* 2001; 412(1):21–5. [https://doi.org/10.1016/s0014-2999\(00\)00923-7](https://doi.org/10.1016/s0014-2999(00)00923-7) PMID: 11166732
101. Nilsson S, Mäkelä S, Treuter E, Tujague M, Thomsen J, Andersson G, et al. Mechanisms of estrogen action. *Physiol Rev.* 2001; 81(4):1535–65. <https://doi.org/10.1152/physrev.2001.81.4.1535> PMID: 11581496
102. Wilson ME, Dimayuga FO, Reed JL, Curry TE, Anderson CF, Nath A, et al. Immune modulation by estrogens: role in CNS HIV-1 infection. *Endocrine.* 2006; 29(2):289–97. <https://doi.org/10.1385/ENDO:29:2:289> PMID: 16785604
103. Asin SN, Heimberg AM, Eszterhas SK, Rollenhagen C, Howell AL. Estradiol and progesterone regulate HIV type 1 replication in peripheral blood cells. *AIDS Res Hum Retroviruses.* 2008; 24(5):701–16. <https://doi.org/10.1089/aid.2007.0108> PMID: 18462082
104. Rodriguez-Garcia M, Biswas N, Patel MV, Barr FD, Crist SG, Ochsenbauer C, et al. Estradiol reduces susceptibility of CD4+ T cells and macrophages to HIV-infection. *PLoS One.* 2013; 8(4):e62069. <https://doi.org/10.1371/journal.pone.0062069> PMID: 23614015
105. Daher A, Longuet M, Dorin D, Bois F, Segéral E, Bannwarth S, et al. Two dimerization domains in the trans-activation response RNA-binding protein (TRBP) individually reverse the protein kinase R inhibition of HIV-1 long terminal repeat expression. *J Biol Chem.* 2001; 276(36):33899–905. <https://doi.org/10.1074/jbc.M103584200> PMID: 11438532
106. Kino T, Slobodskaya O, Pavlakis GN, Chrousos GP. Nuclear receptor coactivator p160 proteins enhance the HIV-1 long terminal repeat promoter by bridging promoter-bound factors and the Tat-P-TEFb complex. *J Biol Chem.* 2002; 277(4):2396–405. <https://doi.org/10.1074/jbc.M106312200> PMID: 11704662
107. Kruczynski A, Barret JM, Etiévant C, Colpaert F, Fahy J, Hill BT. Antimitotic and tubulin-interacting properties of vinflunine, a novel fluorinated Vinca alkaloid. *Biochem Pharmacol.* 1998; 55(5):635–48. [https://doi.org/10.1016/s0006-2952\(97\)00505-4](https://doi.org/10.1016/s0006-2952(97)00505-4) PMID: 9515574
108. Capasso A. Vinorelbine in cancer therapy. *Curr Drug Targets.* 2012; 13(8):1065–71. <https://doi.org/10.2174/138945012802009017> PMID: 22594474
109. Garcia CC, Guabiraba R, Soriani FM, Teixeira MM. The development of anti-inflammatory drugs for infectious diseases. *Discov Med.* 2010; 10(55):479–88. PMID: 21189219
110. Di Petrillo A, Orrù G, Fais A, Fantini MC. Quercetin and its derivatives as antiviral potentials: A comprehensive review. *Phytother Res.* 2022; 36(1):266–78. <https://doi.org/10.1002/ptr.7309> PMID: 34709675
111. Ono K, Nakane H. Mechanisms of inhibition of various cellular DNA and RNA polymerases by several flavonoids. *J Biochem.* 1990; 108(4):609–13. <https://doi.org/10.1093/oxfordjournals.jbchem.a123251> PMID: 2292590
112. Pasetto S, Pardi V, Murata RM. Anti-HIV-1 activity of flavonoid myricetin on HIV-1 infection in a dual-chamber in vitro model. *PLoS One.* 2014; 9(12):e115323. <https://doi.org/10.1371/journal.pone.0115323> PMID: 25546350
113. Zheng Y, Jiang X, Gao F, Song J, Sun J, Wang L, et al. Identification of plant-derived natural products as potential inhibitors of the Mycobacterium tuberculosis proteasome. *BMC Complement Altern Med.* 2014; 14:400. <https://doi.org/10.1186/1472-6882-14-400> PMID: 25315519
114. Villaume SA, Fu J, N'Go I, Liang H, Lou H, Kremer L, et al. Natural and Synthetic Flavonoids as Potent Mycobacterium tuberculosis UGM Inhibitors. *Chemistry.* 2017; 23(43):10423–9. <https://doi.org/10.1002/chem.201701812> PMID: 28497493
115. Bulmer JN, Williams PJ, Lash GE. Immune cells in the placental bed. *Int J Dev Biol.* 2010; 54(2–3):281–94. <https://doi.org/10.1387/ijdb.082763jb> PMID: 19876837
116. Shukla H, Kumar V, Singh AK, Rastogi S, Khan SR, Siddiqi MI, et al. Isocitrate lyase of Mycobacterium tuberculosis is inhibited by quercetin through binding at N-terminus. *Int J Biol Macromol.* 2015; 78:137–41. <https://doi.org/10.1016/j.ijbiomac.2015.04.005> PMID: 25869309