SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •
Special Focus on Biomolecular Network Analysis and Application

July 2016, Vol. 59 070102:1–070102:15
doi: 10.1007/s11432-016-5580-2

Identifying disease modules and components of viral infections based on multi-layer networks

Yuanyuan LI[1,2,3] & Xiufen ZOU[1,2*]

[1]*School of Mathematics and Statistics, Wuhan University, Wuhan* 430072*, China;*
[2]*Computational Science Hubei Key Laboratory, Wuhan University, Wuhan* 430072*, China;*
[3]*School of Science, Wuhan Institute of Technology,Wuhan* 430074*, China*

Received March 31, 2016; accepted April 18, 2016; published online June 7, 2016

Abstract With the emergence of multi-dimensional data, we can more comprehensively analyze the pathogenic mechanisms of complex diseases, and thereby improve the diagnosis, treatment and prevention of these diseases. This study presents a novel multi-layer network-based strategy that integrates multi-dimensional data, and identifies disease-related modules and components of viral infections. We first propose a systematic method that constructs a virus-host interaction network with three layers: a viral protein layer, a host protein layer and a host gene layer. This method combines the data of high-throughput gene expression, viral protein interactions, virus-host interactions, protein-protein interactions and transcriptional regulatory relationships. To extract the underlying mechanisms of viral infections from the multi-layer networks, we identify the intra-layer and cross-layer modules, and investigate the conserved modules across multiple datasets. The essential components in the multi-layer networks are detected by singular-value decomposition. The identified conserved modules and essential components are combined into a functional enrichment analysis that reveals their contributions to influenza virus replication. By this analysis, we elucidate the common and specific mechanisms of the replication cycles of two influenza strains. By combining the different layers of information, we can comprehensively understand pathogenic mechanisms of complex diseases.

Keywords multi-dimensional data, multi-layer network, complex disease, virus infection, disease modules

Citation Li Y Y, Zou X F. Identifying disease modules and components of viral infections based on multi-layer networks. Sci China Inf Sci, 2016, 59(7): 070102, doi: 10.1007/s11432-016-5580-2

1 Introduction

Infectious diseases, especially viral infections such as influenza, human immunodeficiency virus, hepatitis and dengue fever, pose a worldwide threat to public health. These viruses rely on their host's cellular machinery to complete their replication cycle [1]. By understanding the basic replication mechanisms of viruses, we can decipher their pathogenic mechanisms and develop preventive strategies against potential human pandemics.

Using recently developed high-throughput technologies and computational methods, researchers have deduced various biological networks [2–4], such as protein-protein interaction networks, transcriptional regulatory networks and metabolic networks. Network-based approaches allow the characterization and

* Corresponding author (email: xfzou@whu.edu.cn)

Table 1 Five gene expression datasets for two different types of influenza A viruses

GEO accession	Viral agent	Time post infection	Strain/Line
GSE37571	California/04/2009 (H1N1)	0,3,7,12,18,24,30,36 and 48 hpi	Calu-3 cells
GSE40844	California/04/2009 (H1N1)	0,12,24 and 48 hpi	Calu-3 cells
GSE40844	Netherlands/602/09 (H1N1)	0,3,7,12,18,24,30,36 and 48 hpi	Calu-3 cells
GSE19580	Wellington/43/2006 (H3N2)	2,6 and 24 hpi	Calu-3 cells
GSE49840	Panama/2007/99 (H3N2)	3,7,12 and 24 hpi	Calu-3 cells

exploration of diverse biological processes, including the pathogenic mechanisms of viral infections [5]. To date, network-based analysis methods have focused on single biological networks such as those mentioned above. Few studies have combined several kinds of networks. However, the biological relationships described by different networks are usually interdependent, that is, they interact and influence each other through sophisticated mechanisms that maintain the function and robustness of cells [6, 7]. Therefore, studying single networks in isolation may not reveal the molecular mechanisms underlying the biological relationships among multi-layer networks. Accumulating experimental studies have delivered large, heterogeneous datasets at multiple levels. However, a systematic method that can integrate these multi-dimensional data into multi-layer networks presents a new challenge. If such a tool were available, we could unravel the intricate pathogenic mechanisms of viruses.

During the course of a viral infection, viral proteins interact with a group of host proteins and pathways. Viral and host factors and virus-host relationships have been extensively studied over the last decade [8, 9]. Other researchers have integrated protein-protein and protein-DNA interactions to infer signaling-regulatory pathways [10, 11]. However, the protein-protein interactions between virus and host, the protein-gene interactions in the host, and the disease-related modules involved in the virus life cycle, have not been integrated to date. This lack of a systematic integrative approach may limit the effectiveness of network-based approaches in deciphering the pathogenic mechanisms of viruses. To address this problem, we propose a general framework that integrates multi-dimensional data into a multi-layer network. The framework then mines the possible modules involved in the individual steps of viral replication, and uncovers the essential components of viral replication. To verify the framework, we apply it to five influenza virus datasets: three datasets of strain H1N1 and two datasets of strain H3N2. Using an identified atlas of conserved modules and the essential components of H1N1 and H3N2 infection, respectively, we uncover the underlying biological phenomena and gain insights into the annotated biological activities within the cell during influenza A infection.

2 Methods

2.1 Data collection

The gene expression profiling datasets of two influenza strains (H1N1 and H3N2 influenza A virus (IAV)) were downloaded from the NCBI GEO database (GSE37571, GSE40844, GSE19580 and GSE49840) [12]. Ignoring the probe sets without corresponding gene symbols, we averaged the expression values of probe sets mapped to the same gene. To evaluate the efficiency and robustness of our proposed approach, we collected two or three independent datasets for each strain. Table 1 details the information corresponding to each dataset.

In the biological dataset GSE37571, Calu-3 cells were infected with A/CA/04/2009 (H1N1) influenza virus at an MOI (Multiplicity of infection) of 3. Samples were collected at 0, 3, 7, 12, 18, 24, 30, 36 and 48 hours post-infection (hpi). There are 3 mock and 3 infected replicates at each time point.

In the biological dataset GSE40844, cells were infected at an MOI of 3. Samples of A/Netherlands/602 /09 (H1N1)-infected and mock-infected cells were collected at 0, 3, 7, 12, 18, 24, 30, 36, and 48 hpi. Samples of A/California/04/2009 (H1N1)-infected cells were collected at 0, 12, 24, and 48 hpi. All samples were collected in triplicate.

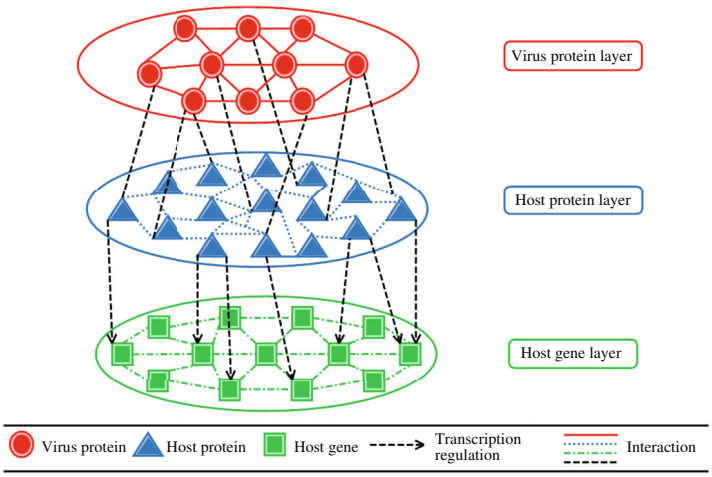

Figure 1 (Color online) Schematic diagram representing three-layer network structure among virus proteins, host proteins and genes.

The biological dataset GSE19580 contains 9 samples of H3N2-infected cells. There is one replicate at 2 hpi, three replicates at 6 and 24 hpi and two replicates of media control samples at 24 hpi [13].

In the biological dataset GSE49840 [14], Calu-3 cells were apically infected with a human seasonal virus A/Panama/2007/1999 (H3N2) at an MOI of 1. There are 4 mock replications and 3 H3N2-infected replications at 3 hpi, and 4 mock and 4 H3N2-infected replications at 7, 12, 24 hpi.

2.2 Construction of multi-layer networks

The multi-layer network is composed of three layers: a viral protein-protein interaction network, a host protein-protein interaction network and a host gene regulatory network. The nodes in each layer are viral proteins, host proteins and host genes, respectively. This three-layer network is schematized in Figure 1. The nodes of the multi-network are the viral protein, the host protein and the host genes, respectively. The interactions among the different three layers and between any two layers are obtained as follows.

2.2.1 *Mining viral protein-protein interaction data and viral protein-host protein association*

Influenza A viruses are segmented, single-stranded, negative-sense RNA viruses encoding at least 10 viral proteins [15]. Some strains encode two additional viral proteins, PB1-F2 and PB1 N40 [16,17]. Using a yeast two-hybrid (Y2H) approach, Shapira et al. [8] systematically identified the direct binary contacts among the ten major viral proteins, and between each viral protein and the 12000 human proteins available in the Human ORFeome v3.1 collection [18]. They discovered 31 intraviral interactions among the 10 viral proteins (NP, PA, PB1, M1, M2, NA, NS2, PB2, HA, NS1), and 135 pairwise interactions between the ten H1N1 viral proteins and 87 human proteins. They also detected 81 interactions between ten H3N2 viral proteins and 66 human proteins.

2.2.2 *Collecting host protein-protein interactions and transcription regulations*

Interactions within the host proteins, which directly interact with viral proteins, interactions between them and other proteins with interaction scores exceeding 0.7, and interactions among the above proteins with interaction scores exceeding 0.7 were downloaded from the Human Integrated Protein-Protein Interaction rEference (HIPPIE) [19]. After filtering out the self- and repeated interactions, the largest connected component of the network was considered as the host protein layer network. Altogether, there are 1579 interactions among 304 host proteins in H1N1-infected cells, and 1555 interactions among 287 host proteins in H3N2-infected cells (see Appendix, Tables A1 and A2 respectively). The protein-protein interaction network and gene regulatory network are interconnected through the transcription factor (TF) interface. The potential TFs and the regulatory associations between the TFs and target genes

(TGs) were obtained from the Integrated Transcription Factor Platform (ITFP) [20, 21]. Since 1998, 1449 human genes in six independent genome-wide screens have been identified as potential host factors in influenza virus replication [9]. Potential regulator-target links were filtered by limiting each node of the incoming links to a potential host gene involved in influenza virus replication. After the filtering, we obtained two initial transcriptional regulatory networks; the first consisting of 52 TF nodes, 193 TG nodes (see Appendix, Table A3) and 399 regulatory links of H1N1 infection, the second containing 52 TF nodes, 200 TG nodes (see Appendix Table A4) and 400 regulatory links of H3N2 infection.

2.2.3 *Optimization of transcriptional regulatory network*

During transcription, a TF bound to the DNA sequence recruits RNA polymerase II to a DNA promoter region, initiating the transcription procedure [22, 23]. Transcription can be simply described by the following linear model:

$$(\mathrm{TG})_i = (\mathrm{TF})\Theta_i, \tag{1}$$

where $(\mathrm{TG})_i = ((\mathrm{TG})_i^1,\ (\mathrm{TG})_i^2, \ldots, (\mathrm{TG})_i^m)^{\mathrm{T}},\ i = 1, 2, \ldots, n$ is a vector, representing the expression level of target gene i $(i = 1, 2, \ldots, n)$, n is the number of target genes, and m is the number of samples. $(\mathrm{TF}) = (\mathrm{tf}_j^k)_{m\times c},\ \mathrm{tf}_j^k$ denotes the activity of transcription factor j in sample k, c is the number of transcription factors. $\Theta_i = (\theta_i^1, \theta_i^2, \ldots, \theta_i^c)^{\mathrm{T}},\ \theta_i^j$ represents the transcription regulation strength of the jth TF to the ith TG. Therefore, $\Theta = \{\theta_i^j,\ i = 1, 2, \ldots, n,\ j = 1, 2, \ldots, c\}$ is a parameter set.

When constructing a transcriptional regulatory network, we must identify the parameters in the model (1). The problem becomes the following optimization problem:

$$\min_{\Theta_i}\{||(\mathrm{TG})_i - (\mathrm{TF})\Theta_i||_2^2 + \gamma||\Theta_i||_1\}. \tag{2}$$

This least-squares problem with a penalty on the L1-norm of the parameters can be efficiently solved by MATLAB's LASSO tool [24]. The regularization parameter γ, which controls the sparseness of the resulting networks, can be optimized by the Bayesian information criterion (BIC) score. Specifically, γ is optimized as follows:

$$\min_{\gamma\in\Gamma} \mathrm{BIC}(\gamma) = \ln\left(\frac{\|(\mathrm{TG})_i - (\mathrm{TF})\Theta_i^\gamma\|_2^2}{n}\right) + \frac{\ln n}{n}\|\Theta_i^\gamma\|_0, \tag{3}$$

$$\Theta_i^\gamma \in \arg\min \frac{1}{2}\|(\mathrm{TG})_i - (\mathrm{TF})\Theta_i\|_2^2 + \frac{\gamma}{2}\|\Theta_i\|_1, \tag{4}$$

where $\Gamma = \{\gamma_0, \gamma_1, \ldots, \gamma_L\},\ \gamma_i = \gamma_0\rho,\ 0 < \rho < 1$, the whole procedure for solving this optimization regularization parameter γ problem is presented as follows:

First, for a given $\gamma \in \Gamma$, the optimization problem (4) is solved to obtain Θ_i^γ. Second, Θ_i^γ is substituted into (3) to obtain a BIC score. Finally, γ is optimized by minimizing BIC.

In the end, the significance of the transcription regulation relationship is tested by the Z-statistic [25]. Firstly, the transcription regulation strengths are normalized by

$$\hat{\theta}_i^j = \frac{(|\theta_i^j| - \min(|\Theta|))}{(\max(|\Theta|) - \min(|\Theta|))},\ \theta_i^j \in \Theta,$$

where $|\cdot|$ is an absolute value, $\min(|\Theta|)$ and $\max(|\Theta|)$ are the minimum and maximum value of absolute value for all transcription regulation strengths, respectively. Secondly, z-transforms of the transcription regulation strengths are calculated, by the following equation:

$$z_i^j = \frac{1}{2}\log\left(\frac{1+\hat{\theta}_i^j}{1-\hat{\theta}_i^j}\right).$$

Lastly, classical decision theory yields the following rule when using the significance level α, we reject the null hypothesis $H_0 : \theta_i^j = 0$, if $\Phi^{-1}(z_i^j) < \alpha$, where $\Phi(\cdot)$ is a normal cumulative distribution function of the standard normal distribution $N(0, 1)$. If H_0 is not rejected, then we consider $\theta_i^j = 0$, and we judge the ith TG and the jth TF as being independent of each other.

2.2.4 *Inference of gene regulatory network*

The gene regulatory network represents the causality of a developmental or regulatory process. The expression profiles of two interacting genes are always temporally related. The mutual information (MI) characterizing the nonlinear dependency between two genes X and Y is computed as [26]

$$\mathrm{MI}(X,Y)=\sum_{x\in X}\sum_{y\in Y}p(x,y)\log\frac{p(x,y)}{p(x)p(y)}. \tag{5}$$

Assuming the widely accepted Gaussian distribution of gene expression data, Eq. (5) reduces to the simplified form [27]:

$$\mathrm{MI}(X,Y)=\frac{1}{2}\log\frac{|C(X)|\cdot|C(Y)|}{|C(X,Y)|}, \tag{6}$$

where C is the covariance matrix of variable, $|C|$ is the determinant of matrix C.

A high MI score indicates a close relationship between two paired genes, whereas a low MI score implies independence between the genes. First, the MIs between all possible gene-gene pairs are computed by (6). If the MI score is below a given threshold, the interaction is regarded as a noisy interaction and deleted from the analysis.

Assuming linear interactions among the genes for simplicity, the gene regulatory network is further simplified by minimizing the error between the simulation results and experimental data:

$$\begin{aligned}&\min_{\Phi_i}\left\{\sum_{k=1}^{m}\left\|x_i^{k(\mathrm{sim})}-x_i^{k(\mathrm{exp})}\right\|_{p1}+\lambda\|\Phi_i\|_{p2}\right\}\\&\text{s.t. }\frac{\mathrm{d}X_i}{\mathrm{d}t}=\sum_{j=1}^{n}\phi_{ij}X_j,\end{aligned} \tag{7}$$

where $\Phi_i=(\phi_{i1},\phi_{i2},\ldots,\phi_{in})^{\mathrm{T}}$, ϕ_{ij} denotes the interaction ability from the jth gene to ith gene, $X_i=(x_i^1,x_i^2,\ldots,x_i^m)^{\mathrm{T}},\ i=1,2,\ldots,n$ represent the expression level of gene, m is the number of samples, n is the number of genes, $x_i^{k(\mathrm{exp})}$ and $x_i^{k(\mathrm{sim})}$ are the experimental and simulation data for the kth sample, respectively, and $\|\cdot\|_{p1}$ and $\|\cdot\|_{p2}$ are 1-norm or 2-norm. Here we set $\|\cdot\|_{p1}$, $\|\cdot\|_{p2}$ to be 1-norm. λ is a regularization parameter that is used to balance the accuracy, sparse in the objective function. The above BIC score method can also be used to optimize λ.

The data containing the desired samples are extracted from the microarray data by cubic spline interpolation. The derivatives can then be substituted by central difference approximations for convenience.

Let $Y_i=\frac{\mathrm{d}X_i}{\mathrm{d}t}$, and the problem (7) is approximately transformed into the following problem:

$$\min_{\Phi_i}\left\{\sum_{k=1}^{m}|y_i^k-\sum_{j=1}^{n}\phi_{ij}x_j^k|+\lambda\sum_{j=1}^{n}|\phi_{ij}|\right\}. \tag{8}$$

Let

$$u_k+v_k=\left|y_i^k-\sum_{j=1}^{n}\phi_{ij}x_j^k\right|,\ u_k-v_k=y_i^k-\sum_{j=1}^{n}\phi_{ij}x_j^k,$$
$$\xi_j+\eta_j=|\phi_{ij}|,\ \xi_j-\eta_j=\phi_{ij},$$

where $k=1,2,\ldots,m,\ j=1,2,\ldots,n,\ u_k,v_k,\xi_j,\eta_j\geqslant 0$. Then, problem (8) can be written as a standard linear programming (LP) problem:

$$\begin{aligned}&\min_{u_k,v_k,\xi_j,\eta_j}\ \sum_{k=1}^{m}(u_k+v_k)+\lambda\sum_{j=1}^{n}(\xi_j+\eta_j)\\&\text{s.t.}\quad u_k-v_k=y_i^k-\sum_{j=1}^{n}(\xi_j-\eta_j)\,x_j^k\\&\qquad u_k,v_k,\xi_j,\eta_j\geqslant 0.\end{aligned} \tag{9}$$

The above LP problem (9) can be solved efficiently by MATLAB's linprog function.

Finally, the significant gene regulation relationship is tested by the Z-statistic, as stated above. These three steps are summarized in Figure 2, which presents a framework of the gene regulatory network.

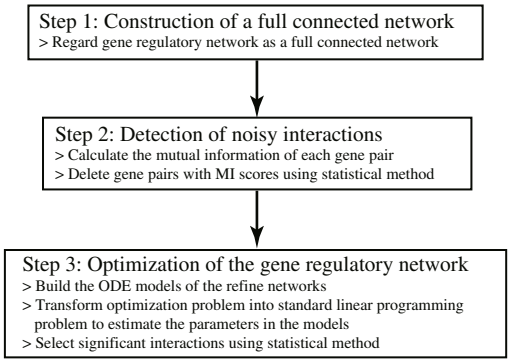

Figure 2 The framework of the gene regulatory network referring.

2.3 Identification of conserved modules

Modules are the building blocks of higher-level functional organizations. An intra-layer and cross-layer module can be defined as a subset of nodes belonging to the same layer and different layers, respectively. First, we detected the intra-layer and cross-layer network modules from the constructed three-layer network using module detection methods, such as the ClusterONE algorithm [28]. Second, to identify the conserved modules across independent distinct datasets of the same influenza strain, we defined the similarity score $\mathrm{SS}(M_i, M_j)$ of pairs of modules M_i and M_j [29]:

$$\mathrm{SS}(M_i, M_j) = \frac{|M_i \cap M_j|}{\min(|M_i|, |M_j|)}, \tag{10}$$

where $|M_i \cap M_j|$ is the number of elements in the intersection of two modules M_i and M_j, and $\min(|M_i|, |M_j|)$ is the minimum of two modules element number [30]. If the similarity of two modules M_i and M_j is equal to or larger than 0.6, we consider them as the same module. If a module appears in different datasets of the same influenza virus type, we call it a conserved module of a influenza virus.

2.4 Functional annotation of modules

The modules were annotated according to the GO Biological Processes (BP), GO Cellular Component (CC), GO Molecular function (MF) and the KEGG pathway annotations of their constituent components. The annotation was performed in the DAVID bioinformatics database [31], and the enrichment significance was determined by the DAVID tool. Significant enrichments in the annotation terms were measured by Fishers' exact test. A module was annotated to a function annotation if the P-value was $\leqslant 0.05$.

2.5 Detection of essential components

The three-layer network is represented in the following block adjacency matrix:

$$\tilde{\boldsymbol{A}} = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} & \mathrm{O} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} & \boldsymbol{A}_{23} \\ \mathrm{O} & \mathrm{O} & \boldsymbol{A}_{33} \end{pmatrix}, \tag{11}$$

where O is a null matrix, and $\boldsymbol{A}_{11}, \boldsymbol{A}_{22}, \boldsymbol{A}_{33}$ represent the intra-layer connections among viral proteins, host proteins and host genes, respectively. $\boldsymbol{A}_{12} = \boldsymbol{A}_{21}^{\mathrm{T}}$ represents the interactions between a viral and a host protein, and $\boldsymbol{A}_{23}$ represents the effective abundance of TFs involved in TG regulation. The components of the three-layer network are ranked according to their essential roles. We calculated the largest eigenvalue and the corresponding eigenvector of the block adjacency matrix. For each component, the essential score (ES) was defined as the modulus of the corresponding eigenvector elements. Based on this score, we can rank the components of each influenza strain. The highest ranked components are expected to be related to influenza virus replication.

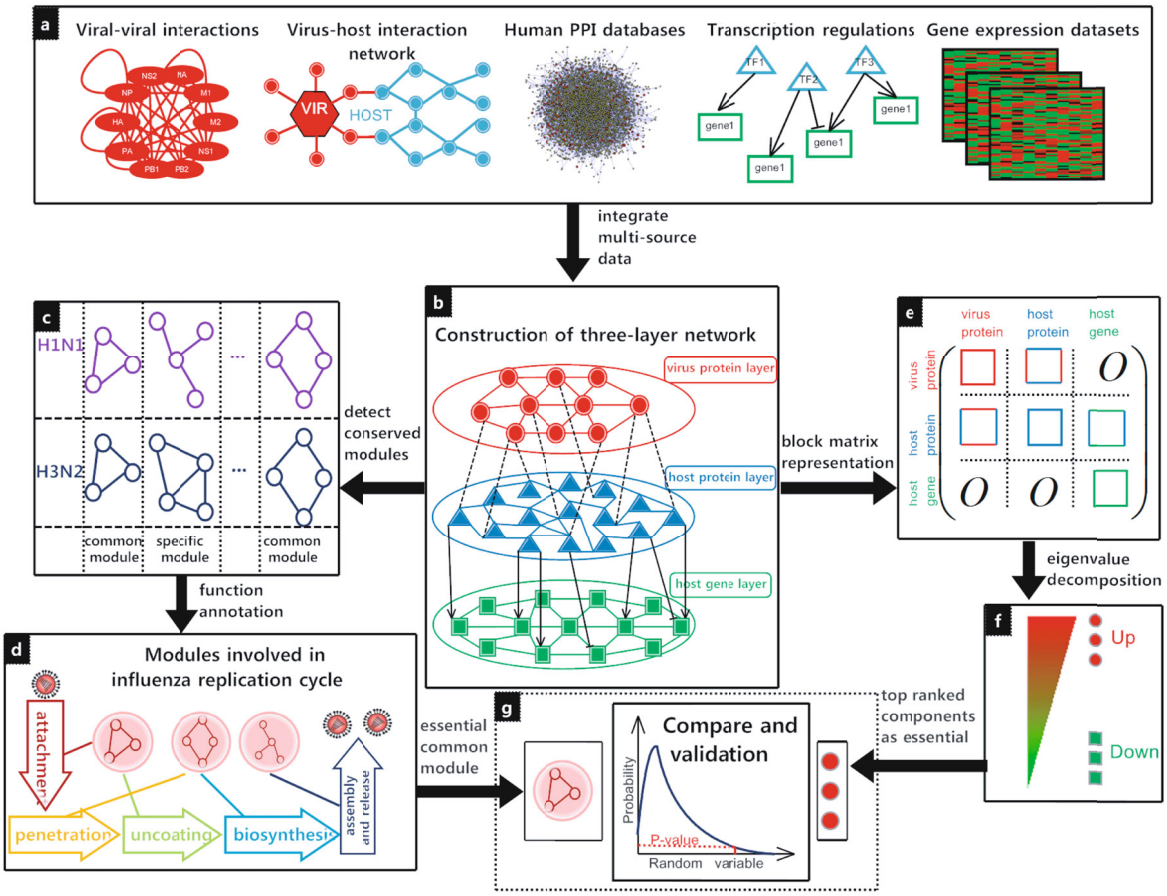

Figure 3 Overview of the proposed paradigm for mining important modules and components for influenza A virus replication. (a) Multi-dimensional data were integrated to construct a multi-layer network. (b) Schematic diagram of the constructed three-layer network. (c) Conserved modules were detected and compared of two different strains of influenza A virus. (d) Functional enrichment analysis of the conserved modules revealed the functional contributions of these modules to the influenza virus replication. (e) A three-layer network was represented by a block matrix. (f) Ranking the nodes of the three-layer network by singular value decomposition method. (g) Overlap analysis of the essential common module and the top ranked components based on inverse cumulative hyper geometric distribution.

3 Results

To identify the important modules and components in influenza A replication, we developed a general paradigm that constructs a multi-layer network and performs system analyses. The proposed paradigm is depicted in Figure 3. We applied our proposed methodology to three H1N1 IAV infection datasets and two H3N2 IAV infection datasets.

3.1 Three-layer network construction

Our general framework constructs a three-layer network of viral protein interactions, virus-host molecular interactions, host protein-protein interactions, and transcriptional regulations with gene expression profiles. The viral protein layer constitutes a viral-viral interaction network based on Shapira et al.'s results [8]. The corresponding physical interactions between the viral and host proteins in H1N1 IAV and H3N2 IAV infections were also extracted from Shapira et al. [8]. The host protein layer contains a PPI network constructed from the PPI database. On the host gene layer, the noisy interactions were first filtered by the MI, which measures the nonlinear dependence between paired nodes. Next, the gene regulatory network was described by a set of ordinary differential equations, and the model parameters were identified by optimization algorithms [29]. Finally, the significance of the gene-gene interactions was determined by a statistical method. The corresponding initial transcriptional regulatory network was

Table 2 Average relative errors for the five data groups

GEO accession	Viral agent	Average relative errors	
		Gene regulation network	Transcriptional regulatory network
GSE37571	California/04/2009 (H1N1)	0.0016	0.0205
GSE40844	California/04/2009 (H1N1)	0.0044	0.0107
GSE40844	Netherlands/602/09 (H1N1)	0.0015	0.0123
GSE19580	Wellington/43/2006 (H3N2)	0.0041	0.0167
GSE49840	Panama/2007/99 (H3N2)	0.0016	0.0276

constructed using the Integrated Transcription Factor Platform. Finally, the transcriptional regulations were depicted in a simplified linear equation model. Again, the model parameters were optimized, and the significances of the transcriptional regulations between TFs and TGs were statistically determined.

Applying the above framework to the three H1N1 IAV datasets and two H3N2 IAV datasets, we constructed five three-layer networks of H1N1 or H3N2 IAV infection.

The accuracies of the constructed transcriptional and gene regulatory networks were evaluated by the average relative error (ARE). In numerical analysis, the relative error is the absolute error divided by the magnitude of the experimental value. Accordingly, the ARE of the transcriptional regulatory network is defined as follows:

$$\mathrm{ARE} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m}\sum_{j=1}^{m}\frac{\left|X_i^{\exp}(t_j) - X_i^{\mathrm{sim}}(t_j)\right|}{|X_i^{\exp}(t_j)|}, \tag{12}$$

$X_i^{\mathrm{sim}}(t_j)$ and $X_i^{\exp}(t_j)$ are the simulation and experimental value of the ith TG at time point t_j, m is the number of samples and n is the number of TGs. $X_i^{\mathrm{sim}}(t_j) = \sum_{k=1}^{c}\theta_i^k \cdot Y_k^{\exp}(t_j)$, where $Y_k^{\exp}(t_j)$ is the experimental value of the kth TF at time point t_j, c is the number of TFs and θ_i^k is the optimal solution of the optimization problem (2), which denotes the transcription regulation strength of the kth TF to the ith TG.

When for gene regulatory network, ARE is defined as below:

$$\mathrm{ARE} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m}\sum_{j=1}^{m}\frac{|X_i^{\exp}(t_j+\Delta t) - X_i^{\mathrm{sim}}(t_j+\Delta t)|}{|X_i^{\exp}(t_j+\Delta t)|}, \tag{13}$$

$X_i^{\mathrm{sim}}(t_j+\Delta t)$ and $X_i^{\exp}(t_j+\Delta t)$ are the simulation and experimental value of the ith host gene at time point $t_j+\Delta t$, m is the number of samples and n is the number of host genes. $X_i^{\mathrm{sim}}(t_j+\Delta t)$ is calculated according to the following formula:

$$X_i^{\mathrm{sim}}(t_j+\Delta t) = X_i^{\exp}(t_j-\Delta t) + 2\cdot\Delta t\cdot\sum_{k=1}^{n}(\phi_{ik}\cdot X_k^{\exp}(t_j)), \tag{14}$$

where ϕ_{ik} is the optimal solution of the optimization problem (7), which denotes the interaction ability from the kth gene to ith gene at time point t_j, Δt stands for the time increment.

Table 2 summarizes the AREs of the gene and transcriptional regulatory networks in H1N1 and H3N2 IAV infection. In all five data groups, the AREs of the gene regulation networks are below 0.005, whereas those of the transcriptional regulatory networks are below 0.03. These results confirm the accuracy of the constructed regulatory networks.

3.2 Identification of modules at each stage of influenza virus replication

Once the virus particles have attached to the surface receptors of their host cells, they can initiate their life cycles. For this purpose, the virus particles enter the cell and uncoat their nucleic acids. The viral ribonucleoproteins (vRNPs) are then transported into the cell nucleus through the nuclear pores. After genome replication, transcription, and protein synthesis, the newly synthesized viral RNA is transported from the nucleus to the cytoplasm. Virion components are transported to the assembly site and assemble into progeny viruses, which then bud from the cells [9].

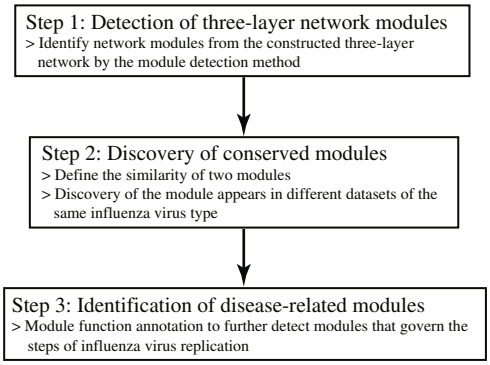

Figure 4 The framework of diseased-related module identification.

Table 3 Enriched module relate functional annotation in each step of influenza virus replication

Module		Functional annotation	P-value		Host cellular biological process	Steps of viral life cycle
H1N1	H3N2		H1N1	H3N2		
M1	M5	Vesicular fraction	4.9×10^{-2}	6.5×10^{-3}	COPI vesicular transport	Endocytosis, fusion and uncoating
		Endocytosis	6.7×10^{-4}	1.7×10^{-2}	Endocytosis	
		Proton acceptor	2.9×10^{-6}	2.3×10^{-5}	V-type ATPase proton transport	
		Post-translational protein modification	6.9×10^{-5}	3.0×10^{-4}	Post-translational modification	vRNP transport from the nucleus to the cytoplasm
		Vesicular fraction	4.9×10^{-2}	6.5×10^{-3}	COPI vesicular transport	Assembly and budding
		Cytoskeleton	3.4×10^{-2}	3.4×10^{-2}	Cytoskeletal components	
M2	M6	Nuclear pore	1.3×10^{-6}	6.5×10^{-8}	Nuclear pore complex	Transport of vRNP into the host nucleus
M3	M7	Nuclear mRNA splicing	4.8×10^{-6}	4.3×10^{-4}	Pre-mRNA splicing	Genome replication/transcription and translation
		Eukaryotic translation initiation factor	2.5×10^{-2}	2.8×10^{-2}	Translation initiation	
		Cytosolic ribosome	1.5×10^{-3}	2.3×10^{-2}	Ribosome	
M4	M8	Post-translational protein modification	5.2×10^{-6}	5.2×10^{-6}	Post-translational modification	vRNP transport from the nucleus to the cytoplasm

To more comprehensively characterize the host machinery exploited by the influenza virus, we employed the ClusterONE algorithm [28] to detect the intra-layer and cross-layer modules in the three-layer networks of each data group. The conserved modules common to all H1N1 IVA infection networks or to all H3N2 IVA infection networks were obtained by defining the similarity between two modules.

The three GO categories BP, MF, and CC were separately analyzed to infer the biological activity of all conserved modules. Our enrichment analysis identified significantly enriched function annotations (P-value < 0.05) in each conserved module.

To understand the modules that govern the steps of influenza virus replication, we assigned the conserved modules to different steps of the influenza virus replication. This assignment was based on the likely host cellular processes in the viral life cycle. The computational framework of the diseased-related module identification is shown in Figure 4. Tables 3 and 4 list the conserved modules that govern the steps of influenza virus replication and the corresponding module components, respectively.

Using (10), we computed the similarity scores between modules M1 and M5, modules M2 and M6, modules M3 and M7, and modules M4 and M8. If the similarity score between the paired modules exceeded 0.6, it was considered as a common module in different influenza infections; otherwise, it was

Table 4 The identified conserved modules that govern the steps of influenza virus replication. The host genes were noted in bold and italic

Subtypes	Modules	Size	Components in the corresponding module
H1N1	M1	49	{TUBA3D TANK CHUK TUBB4A LRRK2 TICAM1 EGFR ***TUBB*** TBK1 SLC25A5 BIRC2 TNFRSF1A DNAJA1 HSPA1A HSPA1L HSPA8 HSPA9 HSP90AA1 IKBKB BIRC3 MAP3K1 NAP1L1 NGFR TNFRSF1B CD40 PARK2 RPS3 SNCA SRC ACTG1 TNFAIP3 TUBA1A TRAF1 TRAF2 TRAF3 TRAF6 TUBA3C YWHAB ARRB1 CASP8 TRADD RIPK1 RIPK2 TNFRSF11A MAP3K14 HGS TIMM50 BAG2 MAP3K5}
	M2	23	{KPNA6 KPNB1 KPNA3 MEPCE MYD88 DOCK8 SPOP DOCK7 PRPF3 TRIM13 ***RPL35 CIRBP CSE1L ERC1 MAZ MINK1 CDK17 SMU1 SAV1 TGFBR1 SMC1A CASP3 NUPL1***}
	M3	44	{PDLIM4 ***RBM12 RBM5 CALCOCO2 SF3A3 KIF2C COPA DFFB E2F1 AHCY MDN1 PRPF31 GTF3C2 KIF11 MAZ MYBL2 NFYA MINK1 CDK12 GTPBP2 PSMD12 PTBP1 ARID1B SAV1 RPS10 CX3CL1 SGK1 SHC1 TGFBR1 TOP2A CAD C6orf62 SMC1A ZNF414 RUVBL1 EIF3B EIF3C EFTUD2*** EXOSC8 EIF4G1 BLM ***SUMO2*** EXOSC9 PRPF3}
	M4	33	{TRIM28 ESR1 PARP1 ZNF451 HDAC1 ZMYND11 HNRNPK PIAS3 AR JUN ETS1 MDM2 HIPK2 MYC PIN1 PML ***TRIM27*** BLM SIAH2 SUMO3 ***SUMO2*** SP100 ***NCOR1*** BRCA1 ***STAT3 TP53*** UBE2I SUMO1 USP7 SNIP1 AXIN1 BHLHE40 PIAS1 CASP8AP2}
H3N2	M5	51	{TUBA3D TANK CHUK TUBB4A LRRK2 TICAM1 MAP3K14 AGO2 MAP3K1 TBK1 SLC25A5 BIRC2 BIRC3 DNAJA1 HSPA1A HSPA1B HSPA1L HSPA4 HSPA8 HSPA9 IARS IKBKB ARRB1 CASP8 RIPK1 MAP3K5 NAP1L1 PARK2 QARS RELB RPS3 SNCA ACTG1 TNFAIP3 EGFR TRADD RIPK2 TRAF1 TRAF2 TRAF3 TRAF6 TUBA3C ***TUBB*** TNFRSF1A TNFRSF1B HSP90AA1 TUBA1A GAPDH}
	M6	17	{KPNA6 KPNB1 KPNA3 MEPCE MYD88 SKP2 DOCK8 SPOP PRPF3 BRCA1 ***CFLAR CHAF1A CSE1L E2F1 KIF11 POLA1 NUP107***}
	M7	36	{***RBM12 RBM5 PAPOLA CREB1 DENND1B EPRS MDN1 FUS FBXO22 KIF11 MYBL2 ATF1 GINS2 NXT2 RELA SAV1 BID FBXL17 SNRPD3 SRP54 TOP2A TP53 TPT1 C6orf62 SMC1A TRRAP RUVBL1 RABEP1*** EXOSC8 EXOSC9 AICDA EIF4G1 EIF3E HMBOX1 BLM ***SUMO2***}
	M8	32	{TRIM28 ESR1 PARP1 ELK1 HDAC1 ZMYND11 HNRNPK PIAS3 HSPB1 AR JUN EIF4G1 HIPK2 MYC PIAS4 PML RANBP2 ***TRIM27*** SIAH2 ***SUMO2*** BRCA1 ***STAT1 STAT3 TP53*** UBE2I SUMO1 ZBTB16 SNIP1 BHLHE40 PIAS1 PIAS2 ***NCOR1*** CASP8AP2}

considered specific to one influenza infection. Table 5 lists the similarity scores of the above module pairs. Most of the common modules launched their effect in the cytoplasm, whereas modules specific to a particular virus tended to operate in the cell nucleus. Modules M1 and M5 were considered as the most important module pair, because of their important role in three steps of the viral life cycle.

3.3 Essential components centered on essential common modules

The 507 components in the H1N1 viral replication network, and the 497 components in the H3N2 viral replication network, were ranked by their corresponding eigenvector elements of the maximum eigenvalue. The top 2%, 4%, 6% and 8% of the ranked components were selected as candidates of essential components involved in the influenza virus life cycle. The top 2% essential components of H1N1 and H3N2 infection, predicted by singular value decomposition (SVD), are listed in Table 6.

Among these essential components, Cao et al. [32] reported that influenza A virus co-opts DNAJA1 to enter the nucleus and enhance its RNA polymerase activity in an Hsp70 cochaperone-independent manner, Momose et al. [33] identified HSP90 as a host factor that stimulates viral RNA polymerase activity. NP-TRF1, PB2-TRF1, PB2-TRF2, NS1-UBE2I and PB1-UBE2I interactions in H1N1 virus infections, and PB2-TRF1, PB2-TRF2 and NS1-UBE2I interactions in H3N2 virus infections, were de-

Table 5 Common modules and specific modules in each step of influenza virus replication

Module		Size of overlap	Module similarity score	Module type	Steps of viral life cycle	Work location
H1N1 (Size)	H3N2 (Size)					
M1(49)	M5(51)	44	0.898	Common module	Endocytosis,fusion and uncoating	Cytoplasm
					vRNP transport from the nucleus to the cytoplasm	
					Assembly and budding	
M2(23)	M6(17)	10	0.588	Specific module	Transport of vRNP into the host nucleus	Nucleus
M3(44)	M7(36)	15	0.417	Specific module	Genome replication/transcription and translation	Nucleus
M4(33)	M8(32)	25	0.781	Common module	vRNP transport from the nucleus to the cytoplasm	Cytoplasm

Table 6 List of top 2% essential components of H1N1 or H3N2 infected predicted by SVD

Symbol	Descriptions	Rank (H1N1)	Rank (H3N2)
HSP90AA1	Heat shock protein 90kDa alpha (cytosolic), class A member 1	3	3
HSPA1A	Heat shock 70kDa protein 1A	4	9
HSPA1B	Heat shock 70kDa protein 1B	–	10
HSPA8	Heat shock 70kDa protein 8	6	4
MAP3K14	Mitogen-activated protein kinase kinase kinase 14	9	6
SUMO2	Small ubiquitin-like modifier 2	8	11
TNFRSF1A	Tumor necrosis factor receptor superfamily, member 1A	11	8
TP53	Tumor protein p53	5	7
TRAF1	TNF receptor-associated factor 1	2	2
TRAF2	TNF receptor-associated factor 2	1	1
TRAF6	TNF receptor-associated factor 6	10	5
TUBB	Tubulin, beta class I	7	14

tected by Shapira et al. [8]. Many of the top 8% essential components are involved in multiple signaling pathways. TRAF2, TNFRSF1A, RIPK1, TP53, IKBKB, MAP3K14, CHUK and TRADD are enriched in apoptosis regulation, EGFR, TRAF2, RELB, TP53, HSPA1A, HSPA1B, HSPA1L, TNFRSF1A, ARRB1, MAP3K1, TRAF6, IKBKB, MAP3K14, MYC, CHUK and HSPA8 are enriched in the MAPK signaling pathway and TRAF2, TBK1, MAP3K1, RIPK1, TRAF6, IKBKB, CHUK, TANK, TRAF3 and TRADD are enriched in the RIG-I-like receptor signaling pathway, which senses cytoplasmic viral RNA.

Furthermore, to deduce the relationship between the candidate essential components and the essential modules M1 and M5, we computed their overlaps and their significance by the inverse cumulative hypegeometric distribution. The P-value is given by

$$P(\text{at least } x|N,m,k) = \sum_{i=x}^{\min(k,m)} P(i|N,k,m) = 1 - \sum_{i=0}^{x-1} P(i|N,k,m), \quad P(i|N,m,k) = \frac{C_m^i C_{N-m}^{k-i}}{C_N^k}, \quad (15)$$

where N is the total number of the elements in the virus replication network, m is the number of candidate essential elements, k is the size of the essential module, and i is the number of common elements between the candidate essential elements and the essential module M1 or module M5.

The results are listed in Table 7. The P-value defines the probability to find at least the number of overlapping components in a random selection of the size of the essential module. The smaller the P-value, the higher the significance of the results.

Table 7 The relationship between candidate essential components and essential module

Size of essential module		Percentage of top essential components (%)	Overlap/Top components		P-value	
H1N1 (M1)	H3N2 (M5)		H1N1	H3N2	H1N1	H3N2
49	51	2	8/10	9/10	2.77×10^{-9}	5.52×10^{-11}
		4	16/20	16/20	2.22×10^{-16}	8.88×10^{-16}
		6	24/30	24/30	0	0
		8	28/41	30/40	0	0

4 Discussion and conclusion

Identifying the disease-related components and modules in a multi-layer network is crucial to elucidating the underlying pathogenic mechanisms of complex diseases and developing effective treatments. This study proposes a systematic strategy that integrates multi-dimensional data into multi-layer networks. It also discovers both common and specific mechanisms of viral infections by identifying multi-layer modules and components. Our work makes three major contributions to systems biology. (1) First, we developed a systematic approach that constructs a multi-layer network from integrated multi-dimensional data. The multi-layer network better represents the actual biological network than single-layer networks, and is more predictive. (2) Applying our approach to H1N1 and H3N2 influenza viruses and referring to gene ontology information, we identified essential disease-related modules at each step of the viral replication cycle. A disease-related module is a connected subgraph containing all molecular determinants of a disease, and contributes to a systematic understanding of the molecular mechanisms underlying that disease [34]. Through comparison and analysis, we deduced that common and specific modules usually function in the cytoplasm and cell nucleus, respectively. (3) We detected disease-related components of viral infections by applying singular value decomposition. Determining the essential components in a multi-layer network is important, because such components are usually critical to cell survival, or provide potential drug targets in the signaling network of a complex disease [35]. Strikingly, in comparison studies, we discovered that essential disease-related components are concentrated in the essential common module. Overall, our study not only offers a novel general framework for revealing the pathogenic mechanisms of viral infections, but can assist the development of new prevention and treatment strategies against viral infections. In future work, we will focus on mathematical modeling and control of multi-layer networks based on disease-related multi-omics data [36–40].

Acknowledgements This work was supported by Major Research Plan of the National Natural Science Foundation of China (Grant Nos. 91530320, 91230118), and Research Foundation of Hubei Province Department of Education (Grant No. Q20151505).

Conflict of interest The authors declare that they have no conflict of interest.

References

1 Hsu N Y, Ilnytska O, Belov G, et al. Viral reorganization of the secretory pathway generates distinct organelles for RNA replication. Cell, 2010, 141: 799–811
2 Jin S Q, Zou X F. Construction of the influenza A virus infection-induced cell-specific inflammatory regulatory network based on mutual information and optimization. BMC Syst Biol, 2013, 7: 105
3 Jin S Q, Li Y Y, Pan R G, et al. Characterizing and controlling the inflammatory network during influenza A virus infection. Sci Rep, 2014, 4: 3799
4 Li Y Y, Jin S Q, Lei L, et al. Deciphering deterioration mechanisms of complex diseases based on the construction of dynamic networks and systems analysis. Sci Rep, 2015, 5: 9283
5 Barabasi A L, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nat Rev Genet, 2011, 12: 56-68
6 Cantini L, Medico E, Fortunato S, et al. Detection of gene communities in multi-networks reveals cancer drivers. Sci Rep, 2015, 5: 17386

7 Li W, Dai C, Liu C C, et al. Algorithm to identify frequent coupled modules from two-layered network series: application to study transcription and splicing coupling. J Comput Biol, 2012, 19: 710–730
8 Shapira S D, Gat-Viks I, Shum B O, et al. A physical and regulatory map of host-influenza interactions reveals pathways in H1N1 infection. Cell, 2009, 139: 1255–1267
9 Watanabe T, Watanabe S, Kawaoka Y. Cellular networks involved in the influenza virus life cycle. Cell Host Microbe, 2010, 7: 427–439
10 Wang Y C, Chen B S. Integrated cellular network of transcription regulations and protein-protein interactions. BMC Syst Biol, 2010, 4: 20
11 Srivastava A, Kumar S, Ramaswamy R. Two-layer modular analysis of gene and protein networks in breast cancer. BMC Syst Biol, 2014, 8: 81
12 Barrett T, Suzek T O, Troup D B, et al. NCBI GEO: mining millions of expression profiles—database and tools. Nucl Acids Res, 2005, 33: D562–566
13 Hsu A C Y, Barr I, Hansbro P M, et al. Human influenza is more effective than avian influenza at antiviral suppression in airway cells. Amer J Resp Cell Mol Biol, 2011, 44: 906–913
14 Josset L, Zeng H, Kelly S M, et al. Transcriptomic characterization of the novel avian-origin influenza A (H7N9) virus: specific host response and responses intermediate between avian (H5N1 and H7N7) and human (H3N2) viruses and implications for treatment options. mBio, 2014, 5: e01102–01113
15 Ozawa M, Kawaoka Y. Taming influenza viruses. Virus Res, 2011, 162: 8–11
16 Chen W, Calvo P A, Malide D, et al. A novel influenza A virus mitochondrial protein that induces cell death. Nat Med, 2001, 7: 1306–1312
17 Wise H M, Foeglein A, Sun J, et al. A complicated message: identification of a novel PB1-related protein translated from influenza A virus segment 2 mRNA. J Virol, 2009, 83: 8021–8031
18 Lamesch P, Li N, Milstein S, et al. hORFeome v3.1: a resource of human open reading frames representing over 10,000 human genes. Genomics, 2007, 89: 307–315
19 Schaefer M H, Fontaine J F, Vinayagam A, et al. HIPPIE: integrating protein interaction networks with experiment based quality scores. PLoS ONE, 2012, 7: e31826
20 Zheng G, Qian Z, Yang Q, et al. The combination approach of SVM and ECOC for powerful identification and classification of transcription factor. BMC Bioinform, 2008, 9: 282
21 Zheng G, Tu K, Yang Q, et al. ITFP: an integrated platform of mammalian transcription factors. Bioinformatics, 2008, 24: 2416–2417
22 Sun N, Carroll R J, Zhao H. Bayesian error analysis model for reconstructing transcriptional regulatory networks. Proc Nat Acad Sci USA, 2006, 103: 7988–7993
23 Wang R S, Jin G, Zhang X S, et al. Modeling post-transcriptional regulation activity of small non-coding RNAs in Escherichia coli. BMC Bioinform, 2009, 10: S6
24 Geeven G, van Kesteren R E, Smit A B, et al. Identification of context-specific gene regulatory networks with GEMULA-gene expression modeling using LAsso. Bioinformatics, 2012, 28: 214–221
25 Saito S, Hirokawa T, Horimoto K. Discovery of chemical compound groups with common structures by a network analysis approach (affinity prediction method). J Chem Inf Model, 2011, 51: 61–68
26 Brunel H, Gallardo-Chacon J J, Buil A, et al. MISS: a non-linear methodology based on mutual information for genetic association studies in both population and sib-pairs analysis. Bioinformatics, 2010, 26: 1811–1818
27 Zhang X, Zhao X M, He K, et al. Inferring gene regulatory networks from gene expression data by path consistency algorithm based on conditional mutual information. Bioinformatics, 2012, 28: 98–104
28 Nepusz T, Yu H, Paccanaro A. Detecting overlapping protein complexes in protein-protein interaction networks. Nat Methods, 2012, 9: 471–472
29 Xiao X Y, Zhang W, Zou X F. A new asynchronous parallel algorithm for inferring large-scale gene regulatory networks. PLoS ONE, 2015, 10: e0119294
30 Zhang W, Zou X F. A new method for detecting protein complexes based on the three node cliques. IEEE/ACM Trans Comput Biol Bioinform, 2015, 12: 879–886
31 Huang D W, Sherman B T, Lempicki R A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc, 2009, 4: 44–57
32 Cao M M, Wei C D, Zhao L L, et al. DnaJA1/Hsp40 is co-opted by influenza A virus to enhance its viral RNA polymerase activity. J Virol, 2014, 88: 14078–14089
33 Momose F, Naito T, Yano K, et al. Identification of Hsp90 as a stimulatory host factor involved in influenza virus RNA synthesis. J Biol Chem, 2002, 277: 45306–45314
34 Menche J, Sharma A, Kitsak M, et al. Uncovering disease-disease relationships through the incomplete interactome. Science, 2015, 347: 1257601
35 de Domenico M, Sole-Ribalta A, Omodei E, et al. Ranking in interconnected multilayer networks reveals versatile nodes. Nat Commun, 2015, 6: 6868
36 Tan J Y, Zou X F. Complex dynamical analysis of a coupled network from innate immune responses. Int J Bifurcat Chaos, 2013, 23: 1350180
37 Tan J Y, Zou X F. Optimal control strategy for abnormal innate immune response. Comput Math Methods Med, 2015, 2015: 386235
38 Wang D J, Jin S Q, Wu F X, et al. Estimation of control energy and control strategies for complex networks. Adv Complex Syst, 2015, 18: 1550018

39 Zou X F, Niu L L, Jin S Q. The mathematical modeling and analysis for S1PR1-mediated cytokine signaling pathway. J Jiangxi Norm Univ (Nat Sci Ed), 2015, 39: 7–14
40 Jin S Q, Niu L L, Wang G, et al. Mathematical modeling and nonlinear dynamical analysis of cell growth in response to antibiotics. Int J Bifurcat Chaos, 2015, 25: 1540007

Appendix A

Table A1 The 304 host proteins selected of H1N1 virus infection

Symbol
ABLIM1 CBS ENKD1 HSPA9 MEPCE PXN SUMO2 TRAF3 ACOT9 CBX5 ESR1 HSPB1 MLH1 QARS SUMO3 TRAF4 ACTG1 CCAR2 ETS1 HSPB2 MYC QTRT1 SUV39H2 TRAF6 AFF4 CCDC172 EWSR1 HTT MYD88 RABGEF1 TACC1 TRAIP AGO2 CCDC85B EXOC3-AS1 IARS NAP1L1 RAD54B TAL1 TRIM13 AIMP2 CCNK EXOSC8 ICT1 NCAPH2 RBPMS TANK TRIM27 AMOT CD27 EXOSC9 IKBKB NCK2 RHOXF2 TARBP2 TRIM28 AP1M1 CD40 FAM107A IKZF1 NCOR1 RIBC2 TBK1 TRIM29 AP2M1 CDC42 FANCG IKZF3 NDC80 RIPK1 TCEA2 TRIM33 APIP CDC42EP4 FATE1 ILF3 NDUFS3 RIPK2 TCEANC TRIM37 AR CEP70 FBF1 ILK NECAB2 RNF5 TCEB3B TRIP6 ARHGEF39 CFLAR FBXO7 JUN NGFR RPS3 TCF12 TSC22D4 ARRB1 CHMP4B FHL3 KCNRG NIF3L1 RUVBL1 TCF3 TUBA1A ATF2 CHMP6 FMR1 KDM1A NONO RUVBL2 TDRD7 TUBA3C ATN1 CHUK FN1 KHDRBS3 NRF1 SDCBP2 TFCP2 TUBA3D ATXN1 CMTM5 FTH1 KPNA3 NRIP1 SEPT1 THAP1 TUBB AURKB COL4A3BP FXR2 KPNA6 NT5DC2 SH3GL3 TICAM1 TUBB4A AXIN1 CREB3 GABPB1 KPNB1 NUDT18 SIAH1 TIFA UBE2D2 BAG2 CRYAB GAPDH KRT15 NUMB SIAH2 TIMM50 UBE2D3 BAG6 CSNK2A1 GCN1L1 LDOC1 PARK2 SLC25A5 TNFAIP3 UBE2I BAHD1 CTNNB1 GMCL1 LMO3 PARP1 SLC9A3R2 TNFRSF11A UROS BANP DAB2 GORASP2 LNX1 PCBD1 SNAP23 TNFRSF12A USHBP1 BCAP29 DAZAP2 HCFC1 LNX2 PDLIM4 SNCA TNFRSF14 USO1 BCAP31 DHX9 HDAC1 LRPPRC PEG3 SNF8 TNFRSF17 USP7 BCL2L1 DICER1 HDAC4 LRRK2 PIAS1 SNIP1 TNFRSF18 VANGL1 BET1 DNAAF5 HGS LTBR PIAS3 SORBS3 TNFRSF19 VCAM1 BHLHE40 DNAJA1 HIPK2 LYPLA1 PIN1 SP100 TNFRSF1A WT1 BIRC2 DOCK7 HIVEP3 LZTS2 PML SPOP TNFRSF1B XIAP BIRC3 DOCK8 HNRNPC MAGEA11 PNMA1 SPTAN1 TNFRSF4 YWHAB BLM DPPA2 HNRNPK MAGEA6 POM121 SRC TNFRSF8 ZBTB16 BLZF1 DVL1 HOOK1 MAGED1 PPM1A SRPK2 TNFRSF9 ZBTB25 BRCA1 DVL2 HOOK2 MAP3K1 PPP2R1A SRPK3 TNFSF9 ZBTB8A C16orf45 DVL3 HOXA1 MAP3K14 PPP2R5C SSBP2 TNNT1 ZMAT3 C1orf94 DYNLL2 HOXA9 MAP3K5 PRKAB2 SSSCA1 TOLLIP ZMAT4 CALCOCO1 EGFR HSP90AA1 MAPK1IP1L PRKDC STAT3 TP53 ZMYND11 CALCOCO2 EIF2AK2 HSPA1A MAPK9 PRKRA STAU1 TRADD ZNF346 CASP8 EIF4G1 HSPA1L MCRS1 PRPF3 STX5 TRAF1 ZNF451 CASP8AP2 ELK1 HSPA8 MDM2 PSMF1 SUMO1 TRAF2 ZNF646

Table A2 The 287 host proteins selected of H3N2 virus infection

Symbol
ABLIM1 CCDC85B EXOSC9 HSPB2 MYD88 RABGEF1 TANK TRAIP ACTG1 CCNK FAM107A IARS NAP1L1 RAD54B TARBP2 TRIM13 ACTN1 CD27 FANCG KBKB NCK2 RANBP2 TBK1 TRIM27 AFF4 CD40 FATE1 ILF3 NCOR1 RB1 TCEA2 TRIM28 AGO2 CEP70 FBF1 ILK NDC80 RBPMS TCEANC TRIM29 AICDA CFLAR FBXO7 IPO13 NECAB2 RELB TCEB3B TRIM33 AIMP2 CHUK FHL3 JUN NGFR RHOXF2 TCF12 TRIM37 AMOT CREB3 FMR1 KCNRG NIF3L1 RIBC2 TCF3 TRIP6 AP1M1 CRX FTH1 KDM1A NINL RIPK1 TDRD7 TSC22D4 AP2M1 CRYAB FXR2 KHDRBS3 NONO RIPK2 TFCP2 TUBA1A APIP CSNK2A1 GABPB1 KPNA1 NRF1 RPS3 THAP1 TUBA3C AR CTNNB1 GAPDH KPNA3 NRIP1 RUNX1T1 TICAM1 TUBA3D ARHGEF39 DAB2 GCN1L1 KPNA6 NT5DC2 RUVBL1 TIFA TUBB ARRB1 DAZAP2 GMCL1 KPNB1 NUDT18 RUVBL2 TIMM50 TUBB4A ATF2 DHX9 GOLGA2 KRT15 NUMB SDCBP2 TNFAIP3 UBE2D1 ATN1 DICER1 GORASP2 LMO3 PARK2 SIAH1 TNFRSF11A UBE2D2 ATXN1 DNAAF5 HCFC1 LNX1 PARP1 SIAH2 TNFRSF12A UBE2D3 AXIN1 DNAJA1 HDAC1 LNX2 PCBD1 SKP2 TNFRSF14 UBE2I BAG2 DOCK7 HGS LRPPRC PDLIM4 SLC25A5 TNFRSF17 UBE2N BAHD1 DOCK8 HIPK2 LRRK2 PEG3 SLC9A3R2 TNFRSF18 UBR5 BANP DPPA2 HIVEP3 LTBR PIAS1 SNCA TNFRSF19 USHBP1 BHLHE40 DVL1 HMBOX1 LZTS2 PIAS2 SNIP1 TNFRSF1A USO1 BIRC2 DVL2 HNRNPC MAGEA11 PIAS3 SORBS3 TNFRSF1B USP7 BIRC3 DVL3 HNRNPK MAGEA6 PIAS4 SPOP TNFRSF4 VANGL1 BLM DYNLL2 HOOK1 MAGED1 PIN1 SPTAN1 TNFRSF8 WT1 BLZF1 DYRK1A HOOK2 MAP3K1 PML SRC TNFRSF9 XIAP BRCA1 EGFR HOXA1 MAP3K14 PNMA1 SRPK3 TNFSF9 ZBTB16 C16orf45 EIF2AK2 HOXA9 MAP3K5 POM121 SSSCA1 TNNT1 ZBTB25 C1orf94 EIF3E HSP90AA1 MAPK1IP1L PPM1A SSX2IP TOLLIP ZBTB8A CALCOCO1 EIF4G1 HSPA1A MAPK9 PRKAB2 STAT1 TP53 ZMAT3 CALCOCO2 ELK1 HSPA1B MBD4 PRKDC STAT3 TRADD ZMYND11 CASP8 ENKD1 HSPA1L MCRS1 PRKRA STAU1 TRAF1 ZNF346 CASP8AP2 ESR1 HSPA4 MDFI PRMT6 SUMO1 TRAF2 ZNF451 CCAR2 EWSR1 HSPA8 MEPCE PRPF3 SUMO2 TRAF3 ZNF581 CCDC172 EXOC3-AS1 HSPA9 MLH1 PSMF1 SUV39H2 TRAF4 ZNF646 CCDC33 EXOSC8 HSPB1 MYC QARS TACC1 TRAF6

Table A3 The 193 host genes selected of H1N1 virus infection

Symbol
ACAP2 CHAF1A ERC1 LSM2 NTHL1 RACGAP1 SGCA TFDP1 AHCY CIRBP FASN LSM4 NUDCD3 RANBP9 SGK1 TFE3 APOBEC3G CKS1B FBXL17 MAP2K2 NUP107 RAPGEF2 SHC1 TGFBR1 ARID1B CLIP1 FBXL5 MAP2K5 NUPL1 RBM12 SIGMAR1 TIMM17A ASAH1 CLU FBXO22 MAP3K2 NXT2 RBM14 SLC4A7 TLK1 ATF1 COPA FCGR2A MAPK1 PAICS RBM34 SLIRP TMF1 ATP2C1 COPE FUS MAPK14 PAK2 RBM5 SMC1A TOP2A BID COPS6 GABPA MAPKAPK2 PAPOLA RELA SMU1 TP53 BUB1B CPSF4 GCAT MARK3 PBK RNPS1 SNRPB TPT1 BUB3 CREB1 GINS2 MAZ PCMT1 ROCK1 SNRPD2 TRAP1 C16orf72 CRY2 GOPC MCF2L PELI1 RPL35 SNRPD3 TRIM27 C6orf62 CSE1L GTF3C2 MDN1 PIK3C2A RPL5 SNRPF TRMT61A CACTIN CSNK1A1 GTPBP2 MINK1 PLEC RPS10 SNW1 TRRAP CAD CX3CL1 HEATR1 MRPL12 PLK4 RPS2 SPAG5 TUBB CALCOCO2 DAPK3 HMGA1 MRPL42 POLA1 RPS5 SRP54 UBAC2 CASP3 DENND1B HNRNPU MTFR1 POLR2H RPS8 SRSF7 UBE2E1 CBX2 DFFB HSPD1 MYBL2 PRKDC RUVBL1 STAT1 WDR18 CCDC51 E2F1 INPPL1 NAGK PRPF31 RXRB STAT3 XPNPEP1 CCT6A E2F3 IQGAP1 NBN PSMB1 SAE1 STIP1 XPO5 CD58 EFTUD2 IRF3 NCOR1 PSMC3 SAV1 STX5 YWHAZ CDK1 EIF3B KANSL1 NFYA PSMD12 SCD SUMO2 ZFAND3 CDK12 EIF3C KIF11 NHP2L1 PTBP1 SELPLG SUPT6H ZFHX3 CDK17 EIF4A3 KIF2C NLE1 PTPN11 SF3A3 SUV420H1 ZMAT3 CDK4 EPRS LEPR NT5C RABEP1 SF3B3 TCF3 ZNF414 CFLAR

Table A4 The 200 host genes selected of H3N2 virus infection

Symbol
ACAP2 CIRBP ERC1 LSM4 NUP107 RABEP1 SF3B3 TCF3 AHCY CKS1B FASN MAP2K2 NUPL1 RACGAP1 SGCA TFDP1 ARID1B CLIC4 FBXL17 MAP2K5 NXT2 RANBP9 SGK1 TGFBR1 ASAH1 CLIP1 FBXL5 MAP3K2 PAICS RAPGEF2 SHC1 TIMM17A ATF1 CLU FBXO22 MAPK1 PAK2 RBM12 SIGMAR1 TIMM44 ATP2C1 COPA FNTB MAPK14 PAPOLA RBM14 SLC4A7 TLK1 BID COPE FUS MAPKAPK2 PBK RBM34 SLIRP TMF1 BUB1B COPS6 GABPA MARK3 PCMT1 RBM5 SMC1A TOP2A BUB3 CPSF4 GCAT MAZ PELI1 RELA SMU1 TP53 C16orf72 CREB1 GINS2 MCF2L PIK3C2A RHOA SNRPB TPT1 C6orf62 CRY2 GOPC MDN1 PIP4K2B RNPS1 SNRPD2 TRAP1 CACTIN CSE1L GTF3C2 MINK1 PLK4 ROCK1 SNRPD3 TRIM27 CAD CSNK1A1 GTPBP2 MRPL12 POLA1 RPL35 SNRPF TRMT61A CALCOCO2 CX3CL1 HEATR1 MRPL42 POLR2H RPL5 SNW1 TRRAP CASP3 DAPK3 HMGA1 MTFR1 PRKDC RPS10 SPAG5 TUBB CBX2 DBT HNRNPU MYBL2 PRPF31 RPS2 SRP54 UBAC2 CCT6A DENND1B HSPD1 NAGK PRPS1 RPS5 SRSF7 UBE2A CD58 DFFB INPPL1 NBN PSMB1 RPS8 STAT1 UBE2E1 CDK1 E2F1 IQGAP1 NCOR1 PSMB3 RUVBL1 STAT3 WDR18 CDK12 E2F3 IRF3 NFKB1 PSMC3 RXRB STIP1 XPNPEP1 CDK17 EFTUD2 KIAA0430 NFYA PSMD11 SAE1 STX5 XPO5 CDK4 EIF3B KIF11 NHP2L1 PSMD12 SAV1 SUMO2 ZFAND3 CFLAR EIF3C KIF2C NT5C PTBP1 SCD SUPT6H ZFHX3 CHAF1A EIF4A3 LEPR NTHL1 PTPN11 SDE2 SUV420H1 ZMAT3 CHM EPRS LSM2 NUDCD3 RAB5A SF3A3 TAOK1 ZNF414