

METHODOLOGY ARTICLE

Open Access

# Systematic identification of transcriptional and post-transcriptional regulations in human respiratory epithelial cells during influenza A virus infection

Zhi-Ping Liu<sup>1</sup>, Hulin Wu<sup>2</sup>, Jian Zhu<sup>3\*</sup> and Hongyu Miao<sup>2\*</sup>

## Abstract

**Background:** Respiratory epithelial cells are the primary target of influenza virus infection in human. However, the molecular mechanisms of airway epithelial cell responses to viral infection are not fully understood. Revealing genome-wide transcriptional and post-transcriptional regulatory relationships can further advance our understanding of this problem, which motivates the development of novel and more efficient computational methods to simultaneously infer the transcriptional and post-transcriptional regulatory networks.

**Results:** Here we propose a novel framework named SITPR to investigate the interactions among transcription factors (TFs), microRNAs (miRNAs) and target genes. Briefly, a background regulatory network on a genome-wide scale (~23,000 nodes and ~370,000 potential interactions) is constructed from curated knowledge and algorithm predictions, to which the identification of transcriptional and post-transcriptional regulatory relationships is anchored. To reduce the dimension of the associated computing problem down to an affordable size, several topological and data-based approaches are used. Furthermore, we propose the constrained LASSO formulation and combine it with the dynamic Bayesian network (DBN) model to identify the activated regulatory relationships from time-course expression data. Our simulation studies on networks of different sizes suggest that the proposed framework can effectively determine the genuine regulations among TFs, miRNAs and target genes; also, we compare SITPR with several selected state-of-the-art algorithms to further evaluate its performance. By applying the SITPR framework to mRNA and miRNA expression data generated from human lung epithelial A549 cells in response to A/Mexico/InDRE4487/2009 (H1N1) virus infection, we are able to detect the activated transcriptional and post-transcriptional regulatory relationships as well as the significant regulatory motifs.

**Conclusion:** Compared with other representative state-of-the-art algorithms, the proposed SITPR framework can more effectively identify the activated transcriptional and post-transcriptional regulations simultaneously from a given background network. The idea of SITPR is generally applicable to the analysis of gene regulatory networks in human cells. The results obtained for human respiratory epithelial cells suggest the importance of the transcriptional, post-transcriptional regulations as well as their synergies in the innate immune responses against IAV infection.

**Keywords:** Influenza virus infection, Regulatory network in epithelial cells, Dimension reduction, Dynamic Bayesian network, Constrained LASSO

\* Correspondence: [jian\\_zhu@urmc.rochester.edu](mailto:jian_zhu@urmc.rochester.edu); [hongyu\\_miao@urmc.rochester.edu](mailto:hongyu_miao@urmc.rochester.edu)

<sup>3</sup>Department of Microbiology and Immunology, University of Rochester, Rochester, NY 14642, USA

<sup>2</sup>Department of Biostatistics and Computational Biology, University of Rochester, Rochester, NY 14642, USA

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

## Background

Seasonal and pandemic influenza A virus (IAV) continues to be a public health threat and to exert a large economic burden worldwide [1]. IAV RNA segments that encode the hemagglutinin (HA) and neuraminidase (NA) proteins can undergo mutation (antigenic drift) or reassortment (antigenic shift), resulting in new viral strains that humans may lack the heterologous immunity against (e.g., the pandemic H1N1 2009 [2]). In such circumstances, the cell-mediated and humoral immune responses are primary, and a better understanding of the molecular mechanisms of immune responses to IAV infection thus becomes necessary to the development of more effective prevention and treatment strategies. Particularly, the regulations of cell gene expression are of significant importance due to their critical roles in shaping the magnitude and timing of immune responses via the promotion or suppression of the production of various proteins and RNAs (e.g., cytokines [3,4] and miRNAs [5]). Since different cell populations with distinct phenotypes and functions are expected to have different gene regulation profiles, it is desirable to characterize the gene regulations for each cell population as in [6,7]. In this study, the respiratory epithelial cells are of primary interest because they are the first barrier and the main target of IAV infection in human [8]. A number of previous studies have measured and analyzed the gene expression profiles in respiratory epithelial cells during IAV infection [9-12]; however, only a few studies investigated the gene expression regulators like transcription factors [13] or miRNAs [14]. Several studies in other biological disciplines show that TFs and miRNAs cooperatively interact with each other and both of them should be considered in the regulatory network model [15,16]. To our best knowledge, simultaneous identification of both the transcriptional (TF) and post-transcriptional (miRNA) regulations on a genome-wide scale in human respiratory epithelial cells post IAV infection has not been sufficiently addressed.

A number of systems biology approaches have been developed to understand the transcriptional or post-transcriptional regulations [17-19], including several studies particularly for influenza H1N1 virus infection [13,20,21]. For example, Butte and Kohane [22] considered the pair-wise mutual information to screen the gene data associations. Friedman et al. [23] employed the Bayesian network model to capture the conditional independencies between genes. Yeung et al. [24] used the singular value decomposition (SVD) and the robust regression for network reverse engineering, given the sparsity of real large-scale networks. Zhang and Horvath [25] proposed the weighted gene co-expression network analysis, which assigns a continuous connection weight between 0 and 1 to each gene pair. Basso et al. [6] combined the mutual information with the data processing

inequality theory to analyze the gene expression profiles in human B cells. Ordinary differential equation models have also been considered in the previous studies [26], in which the computing efficiency issues were addressed for high-dimensional dynamic systems. However, as evaluated in [27,28], the accuracy of these existing methods are not satisfying, and the reliability of network reverse engineering from gene expression data needs to be further improved. More importantly, two key issues are not sufficiently addressed in the previous studies: first, while both TFs and miRNAs are the major regulators in controlling gene expression [19,29], none of the methods mentioned above simultaneously considered both transcriptional and post-transcriptional regulations such that the inferred interactions could be biased; second, most of the existing approaches are purely data-driven and thus the results could have a high false positive rate [30-32].

In this study, we propose a new framework called SITPR, which stands for Systematic Identification of Transcriptional and Post-transcriptional Regulations, to identify the regulatory relationships by exploiting both curated knowledge and time-course expression data of mRNAs and miRNAs. For this purpose, a background regulatory network is constructed on a genome-wide scale by collecting the experimentally-observed interactions in literature or public databases [33,34] as well as the potential regulatory interactions predicted by representative algorithms for quantifying the sequence-selective binding feasibility between TFs/miRNAs and genes [18,35,36]. Since this background network is not cell type or disease specific, only part of the interactions in this background will be activated in respiratory epithelial cells in response to influenza infection. To identify these activated regulatory interactions, we first employ several dimension reduction approaches, including the community detection algorithm based on network modularity [37], the time delay detection method [38,39], the functional principal component analysis (fPCA) for screening differentially expressed genes [40], and the maximum information coefficient (MIC) [41]. After dividing the large network into much smaller modules, we propose the constrained LASSO method and combine it with the dynamical Bayesian network (DBN) model to detect the activated regulatory relationships from time-course expression data within each module. Our simulation studies suggest that the proposed framework can achieve satisfying performance in terms of sensitivity and specificity for networks of different sizes. We then apply the framework to the real experiment data and analyze the topological and functional features of the regulatory network in respiratory epithelial cells in response to H1N1 influenza virus infection.

## Results and discussion

### Background regulatory network obtained for human

Instead of inferring regulatory network only from expression profiling data, we first build a background regulatory network for human (see Additional file 1: Text S1 and Methods), and then identify the activated regulatory relationships during IAV infection using the mRNA and miRNA time-course expression datasets. We collect the documented regulatory relationships among TFs, miRNAs, and genes in various databases and literature, and also incorporate the potential regulatory relationships between regulators and targets predicted according to the sequence motifs of TFBSs (see Methods). For instance, we retrieve information from the ENCODE project [42] (Text S1) so the majority of the regulators and targets in ENCODE (102 out of 119 TFs, 723 out of 736 miRNAs and 13,607 out of 15,131 target genes) are included in our study. The resulted background regulatory network for human contains 23,079 nodes and 369,277 edges, consisting of 1,456 TFs, 1,904 miRNAs and 19,719 target genes. Among all these edges, 55.6% are experimentally validated interactions, and the rest are predicted regulatory relationships. Note that the regulatory interactions in the constructed background network are not limited to particular cell types or disease conditions; therefore, only part of these interactions will be activated under specific conditions (e.g., in respiratory epithelial cells infected by IAV).

For verification purpose, we calculate several statistical measures of the background regulatory network (Additional file 1: Table TS4), which clearly suggest that the obtained background network is different from a random network. More specifically, the clustering coefficient of our background network is 0.117, which is much higher than that of a random network of a comparable size ( $\sim 1.5 \times 10^{-5}$ ) [43]. Also, the node degrees of our background network are found to satisfy the power-law distribution (Additional file 1: Figure TS3). Fitting the power law  $y = \alpha \cdot x^{-\gamma}$ , where  $y$  denotes the number of nodes and  $x$  denotes the node degree, we obtain  $\gamma = 2.126$ , which is between 2 and 3 and thus suggests our background network is scale-free. This value also indicates that our background network is not random [43,44].

The constructed background network is a high-dimensional network. To make the computing cost affordable, we employ several topology-based and data-based dimension reduction methods to divide the background network into smaller modules. We then identify the activated interactions within each module by fitting the dynamic Bayesian network model to the time-course expression data. To control the false positive rate, we also introduce the novel constrained LASSO formulation into the model fitting procedure.

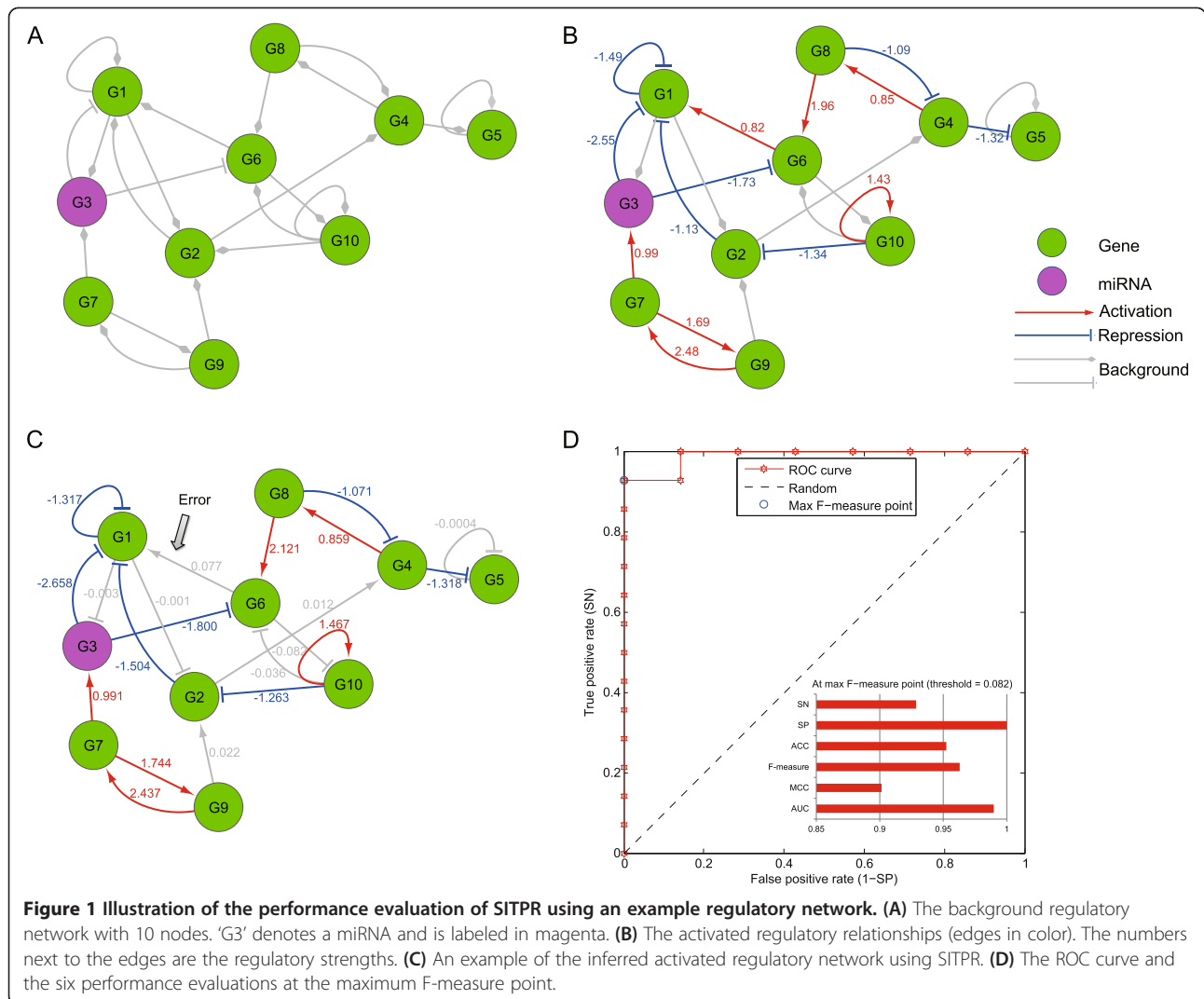
### Algorithm performance evaluation and comparison

Before we apply the SITPR framework to real data, the algorithm performance should be evaluated and compared using networks of different sizes. More specifically, the network size in terms of total node number is 10, 50 or 100, which is chosen to be comparable to the node numbers of the real network modules obtained in this study (ranging from 2 to  $\sim 200$ ). We use linear ordinary differential equations (ODEs) to match the structure of a given network and generate the simulated time-course expression data. To be consistent with the real data used in this study, simulated data are generated at six time points ( $t = 0, \dots, 5$ ) with six replicates at each time point. Gaussian white noises with a standard deviation of 10% of the data mean are added to all the data points. Six commonly-used criteria are chosen to evaluate the algorithm performance, including sensitivity (SN), specificity (SP), accuracy (ACC), F-measure, Matthews correlation coefficient (MCC), and the Area Under ROC Curve (AUC),

$$\begin{aligned}
 SN &= \frac{TP}{TP + FN}, \\
 SP &= \frac{TN}{TN + FP}, \\
 ACC &= \frac{TP + FN}{TP + FP + TN + FN}, \\
 F\text{-measure} &= \frac{2 \times SN \times SP}{SN + SP}, \\
 MCC &= \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FN)(TP + FP)(TN + FP)(TN + FN)}},
 \end{aligned} \tag{1}$$

where TP, FN, FP, and TN are the numbers of true positive, false negative, false positive, and true negative, respectively; also, AUC does not have a closed-form expression and is thus not given in Eq. (1).

For illustration purpose, we show a toy example of the regulatory system that consists of only 10 genes, which has a background regulatory network structure (in gray) as in Figure 1(A) and a set of activated regulatory relationships (in color) as in Figure 1(B). Furthermore, the miRNA is the G3 node labeled in magenta, and the true regulation coefficients are explicitly labeled next to each edge in Figure 1 (B). The inferred regulatory relationships and coefficients from a randomly-chosen simulated dataset are shown in Figure 1(C), and the corresponding evaluation criterion values together with the ROC curve are plotted in Figure 1 (D). Based on 10 simulation runs, we calculate the six criteria and the associated standard deviations. For this toy case, the proposed framework achieves a SN of  $0.886 \pm 0.090$ , SP of  $0.971 \pm 0.069$ , F-measure of  $0.925 \pm 0.070$ , MCC of  $0.875 \pm 0.144$ , ACC of  $0.943 \pm 0.667$ , and AUC of  $0.966 \pm 0.069$  (also see Table 1).



It is necessary to further evaluate the SITPR performance for background regulatory networks of a higher dimension. Since the number of nodes in the largest modules obtained by dividing the real background network (see Methods) is on the order of 100, we consider networks with 50 and 100 nodes. For this purpose, we adopt two network structures from the DREAM3 challenge [28] so both of them have the small-world property and an exponential degree distribution. The smaller structure has 50 nodes and 62 activated regulatory relationships, and the larger structure has 100 nodes and 125 activated regulatory relationships (see Additional file 2, in which 1 denotes a positive regulation and  $-1$  the opposite). We randomly add more inactive edges to the two networks (denoted by 0 in Additional file 2: Table S1) so the smaller background network has 62 out of 82 edges are activated and the larger network has 125 out of 166 edges are activated. Based on such network structures, we use the linear

ODE model again to generate simulated data at six time points with six replicates at each time point, and add 10% Gaussian white noise to each data point.

It is also necessary to compare the proposed framework with other representative or state-of-the-art reverse engineering approaches. For this purpose, we consider the Pearson's correlation coefficient (PCC) method, the mutual information (MI) method [27,28], and the best four algorithms in the DREAM challenge, including CLR [45], ARACNE [46], GENIE3 [47] and TIGRESS [48]. For fairness of comparison, the same task is assigned to all the methods under comparison: identify the activated regulatory relationships from a given background network. The parameter settings of the other methods under comparison are from the corresponding literature, and we make necessary modifications for these methods to take a background network as input.

In Table 1, we summarize the performance comparison results for three different network sizes (10, 50, and 100

**Table 1 Performance evaluation and comparison of SITPR, PCC, MI, CLR, ARACNE, GENIE3 and TIGRESS for networks of three different sizes**

Method	Node Size	SN	SP	ACC	F-measure	MCC	AUC
SITPR	10	0.886 ± 0.090	0.971 ± 0.069	0.943 ± 0.667	0.925 ± 0.070	0.875 ± 0.144	0.966 ± 0.069
	50	0.775 ± 0.089	0.635 ± 0.088	0.670 ± 0.054	0.690 ± 0.039	0.358 ± 0.062	0.703 ± 0.037
	100	0.690 ± 0.080	0.478 ± 0.101	0.530 ± 0.079	0.557 ± 0.080	0.147 ± 0.110	0.573 ± 0.071
PCC	10	0.400 ± 0.241	0.486 ± 0.125	0.457 ± 0.133	0.398 ± 0.204	-0.112 ± 0.293	0.596 ± 0.082
	50	0.495 ± 0.157	0.453 ± 0.167	0.463 ± 0.124	0.444 ± 0.105	-0.043 ± 0.189	0.557 ± 0.038
	100	0.534 ± 0.084	0.526 ± 0.057	0.528 ± 0.045	0.525 ± 0.050	0.052 ± 0.045	0.517 ± 0.027
MI	10	0.243 ± 0.243	0.679 ± 0.097	0.533 ± 0.092	0.293 ± 0.256	-0.101 ± 0.255	0.567 ± 0.062
	50	0.590 ± 0.145	0.527 ± 0.106	0.543 ± 0.065	0.542 ± 0.036	0.103 ± 0.072	0.574 ± 0.027
	100	0.539 ± 0.063	0.569 ± 0.079	0.561 ± 0.050	0.547 ± 0.030	0.095 ± 0.052	0.568 ± 0.024
CLR	10	0.400 ± 0.148	0.543 ± 0.096	0.495 ± 0.068	0.441 ± 0.081	-0.055 ± 0.143	0.540 ± 0.059
	50	0.565 ± 0.133	0.510 ± 0.090	0.523 ± 0.061	0.520 ± 0.072	0.064 ± 0.107	0.539 ± 0.066
	100	0.568 ± 0.130	0.543 ± 0.102	0.549 ± 0.059	0.538 ± 0.044	0.098 ± 0.084	0.550 ± 0.039
ARACNE	10	0.500 ± 0.168	0.643 ± 0.181	0.595 ± 0.117	0.532 ± 0.105	0.151 ± 0.213	0.577 ± 0.084
	50	0.620 ± 0.118	0.508 ± 0.119	0.535 ± 0.067	0.539 ± 0.0275	0.114 ± 0.062	0.547 ± 0.043
	100	0.576 ± 0.112	0.525 ± 0.108	0.537 ± 0.062	0.531 ± 0.039	0.089 ± 0.062	0.554 ± 0.050
GENIE3	10	0.429 ± 0.252	0.521 ± 0.147	0.490 ± 0.163	0.443 ± 0.21	-0.050 ± 0.345	0.609 ± 0.057
	50	0.565 ± 0.116	0.444 ± 0.228	0.473 ± 0.157	0.444 ± 0.162	-0.007 ± 0.187	0.529 ± 0.041
	100	0.512 ± 0.093	0.544 ± 0.062	0.536 ± 0.043	0.520 ± 0.050	0.049 ± 0.077	0.521 ± 0.033
TIGRESS	10	0.629 ± 0.168	0.586 ± 0.171	0.600 ± 0.127	0.583 ± 0.135	0.205 ± 0.233	0.586 ± 0.122
	50	0.490 ± 0.152	0.463 ± 0.105	0.470 ± 0.09	0.461 ± 0.095	-0.041 ± 0.167	0.536 ± 0.067
	100	0.554 ± 0.127	0.507 ± 0.083	0.519 ± 0.058	0.515 ± 0.070	0.053 ± 0.105	0.549 ± 0.033

For fairness of comparison when evaluating the performances, all the methods are assigned the same task (that is, identify the activated regulatory relationships for a given background network).

nodes, respectively) based on 10 simulation runs. For the 10-node networks, the SITPR approach has a remarkable performance in comparison with the other six methods. For example, the AUC of SITPR is  $0.966 \pm 0.069$ , while the PCC, MI, CLR, ARACNE, GENIE3 and TIGRESS methods only achieve an AUC of  $0.596 \pm 0.082$ ,  $0.567 \pm 0.062$ ,  $0.540 \pm 0.059$ ,  $0.577 \pm 0.084$ ,  $0.609 \pm 0.057$ ,  $0.586 \pm 0.122$ , respectively. For the 50-node network, the performance of the SITPR framework is again notably superior to those of the other methods, evidenced by the fact that all the six evaluation criteria of SITPR are the largest among all the approaches under comparison. For instance, the AUC of SITPR is  $0.703 \pm 0.037$ , while the MI and PCC methods have the second and the third best AUCs of  $0.574 \pm 0.027$  and  $0.557 \pm 0.038$ , respectively. As the number of network nodes increase to 100, the performances of all these approaches decrease; however, the SITPR method still outperforms all the other methods in terms of, e.g., AUCs.

To assess the robustness of the SITPR method against data noise, we generate 100 simulated datasets for each of the three noise levels: let the standard deviation of the Gaussian white noise be 10%, 20% or 30% of the data mean. As suggested in Table 2, the performance of the

SITPR method only marginally decreases as the noise level increases, so we conclude that its performance is robust against data noise. For example, for the three different noise levels, the AUCs of our method are  $0.986 \pm 0.045$ ,  $0.924 \pm 0.133$ , and  $0.889 \pm 0.158$ , respectively.

We further evaluate the effects of the total number of edges in the background network on the SITPR performance. In Figure 1, we consider a background regulatory network with 14 activated interactions out of a total of 21 background regulatory relationships. Here we increase the number of background regulatory relationships to 40, 80 and 100, while the number of activated interactions remains at 14. As shown in Table 2, when the number of background regulatory relationships increases, the performance of our approach decreases marginally. Specifically, the AUCs of the SITPR method are  $0.940 \pm 0.060$ ,  $0.893 \pm 0.095$  and  $0.811 \pm 0.071$  for 40, 80 and 100 background regulatory relationships, respectively.

In summary, the simulation study results provide direct evidences for the effectiveness and robustness of the proposed framework for systematically identifying activated transcriptional and post-transcriptional regulations from a given background network.

**Table 2 Effects of the noise level and the total number of background edges on the performance of SITPR based on 100 simulation runs**

Noise	SN	SP	ACC	F-measure	MCC	AUC
10%	0.913 ± 0.086	0.994 ± 0.044	0.967 ± 0.057	0.950 ± 0.057	0.928 ± 0.093	0.986 ± 0.045
20%	0.867 ± 0.143	0.933 ± 0.141	0.911 ± 0.126	0.893 ± 0.131	0.816 ± 0.253	0.924 ± 0.133
30%	0.889 ± 0.182	0.886 ± 0.174	0.887 ± 0.155	0.876 ± 0.171	0.775 ± 0.304	0.889 ± 0.158
Background edge number	SN	SP	ACC	F-measure	MCC	AUC
40	0.884 ± 0.067	0.920 ± 0.071	0.897 ± 0.059	0.900 ± 0.059	0.787 ± 0.120	0.940 ± 0.060
80	0.854 ± 0.108	0.854 ± 0.104	0.854 ± 0.103	0.852 ± 0.100	0.629 ± 0.192	0.893 ± 0.095
100	0.829 ± 0.112	0.770 ± 0.093	0.820 ± 0.095	0.792 ± 0.083	0.521 ± 0.152	0.811 ± 0.071

### Regulations in respiratory epithelial cells during IAV infection

We apply the SITPR framework to understand both the transcriptional and post-transcriptional regulations that are activated in human respiratory epithelial cells (A549 cell line) in response to H1N1 influenza virus infection. In short, from the constructed background regulatory network at the genome-wide scale, we obtain 303 functional modules (2 to ~200 nodes and 1 to ~300 edges in a module) after dimension reduction, and then identify the activated transcriptional and post-transcriptional regulations simultaneously in each module using the dynamic Bayesian network model together with the constrained LASSO formulation (see Methods).

Network motifs are known as the functional building blocks that control gene expressions [15,49]. Therefore, after we determine the activated regulatory interactions from the background network, we further identify 10 types of ‘TF-miRNA-gene’ co-regulation motifs using the FANDOM algorithm [50]. For convenience, the results are summarized in Table 3, in which the numbers of the 10 types of motifs in the background network and in the activated network are both listed. The detailed motif structures can be found in Additional file 1: Figure TS5. The statistical significance of each motif is also evaluated by calculating the associated Z-score. For this purpose, 1000 random networks are generated, and the Z-score is calculated as the difference between the motif occurrence in the real network and its average occurrence in the random networks, divided by the standard deviation of the occurrences in the random networks. We perform such calculations for both the background regulatory network and the activated regulatory network; and a motif with a Z-score greater than 2 is referred as enriched and statistically significant. For respiratory epithelial cells in response to H1N1 influenza virus infection, the ten type of motifs in the background regulatory network after data-based dimension reduction (see Methods) involve 20,310 regulatory relationships among 621 TFs, 642 miRNAs and 7,356 target genes, while nine types of these motifs can be found in the activated network, which involve 4,774

regulatory relationships among 420 TFs, 431 miRNAs and 3,773 target genes (Additional file 3). In these motifs, the numbers of the transcriptional and post-transcriptional regulations are of the same order of magnitude (e.g., 3,140 ‘TF-gene’ versus 1,449 ‘miRNA-gene’ regulations), which suggests that the importance of the post-transcriptional regulatory interactions may be comparable to that of the transcriptional regulatory relationships.

**Table 3 Ten co-regulation motifs and their occurrence frequencies in the background network and the activated network**

Motif ID	Network	Occurrence	Z-Score
M1	Background	1541	24.114
	Active	16	153.470
M2	Background	21317	8.412
	Active	209	2.356
M3	Background	45171	-3.933
	Active	109	-54.482
M4	Background	27987	3.685
	Active	153	1.053
M5	Background	187341	-1.152
	Active	1647	-20.769
M6	Background	2931	-1.091
	Active	-	-
M7	Background	20060	1.017
	Active	59	0.914
M8	Background	3226	-0.624
	Active	13	25.591
M9	Background	481293	-0.478
	Active	38	0.626
M10	Background	88102	-0.230
	Active	273	-24.802

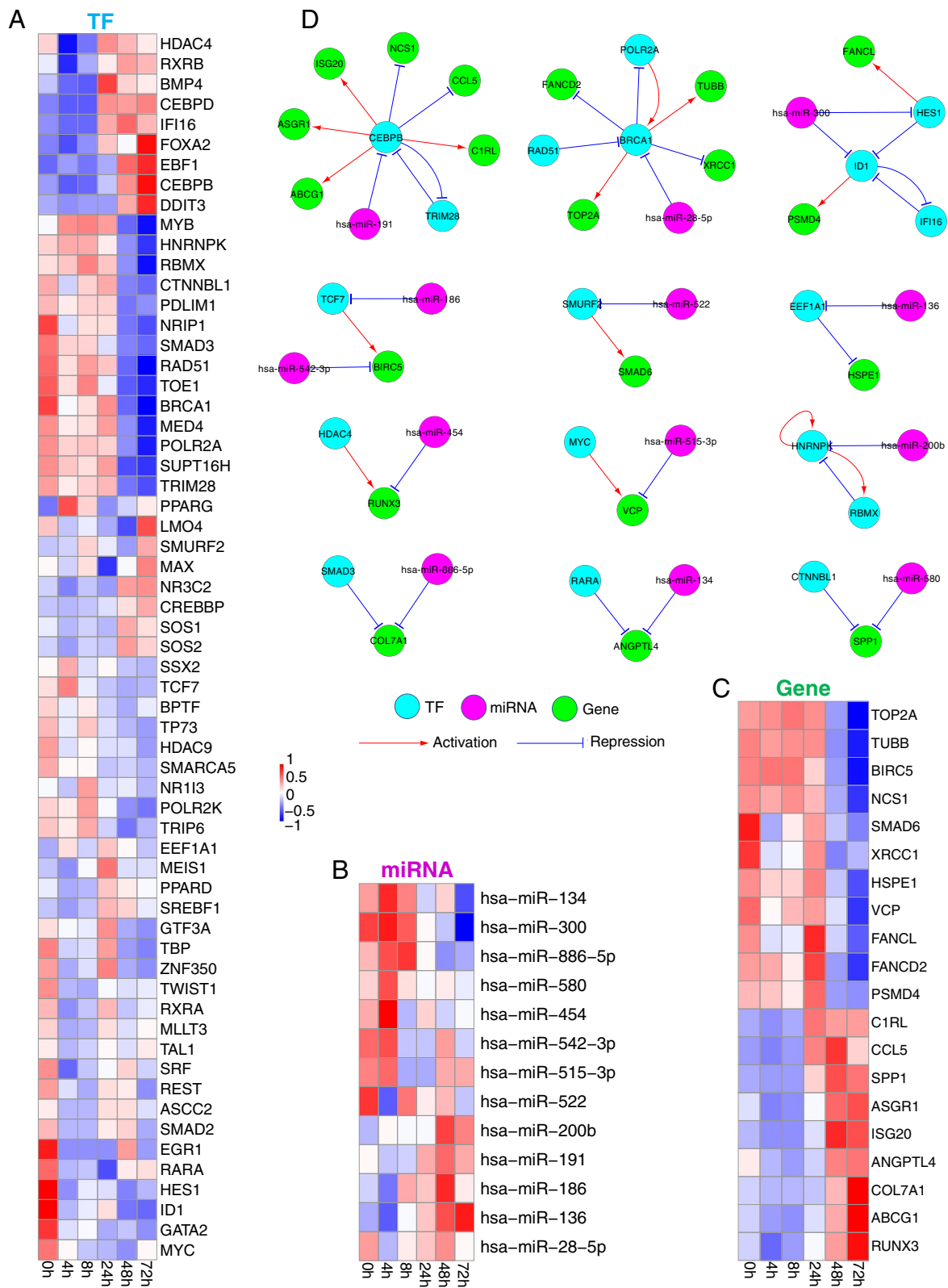
Motifs are ordered according to their absolute Z-scores in the background regulatory network, which are calculated using FANMOD [50]. Motifs ‘M1’, ‘M2’, and ‘M3’ are statistically significant in both the background and the activated regulatory networks (threshold of 2). ‘M5’, ‘M8’ and ‘M10’ are not statistically significant in the background regulatory network, while they become significant in the activated regulatory network.

The motifs in the activated regulatory network (Table 3) are worth of further investigation to better understand the synergy of these motifs in gene expression regulation. We thus combine these co-regulation motifs into larger components if they are connected or overlap with each other. These components involved 61 TFs, 48 miRNAs and 127 genes (Additional file 4). Limited by space, we plot the expression profiles of a part of these TFs, miRNAs, and target genes in Figures 2(A), (B), and (C), respectively; the structures of a part of the components are sketched in Figure 2(D). A specific example is the component centered at 'CEBPB' (CCAAT/Enhancer-Binding Protein Beta), which is an important TF involved in immune and inflammatory responses [51]. In this component (see the upper left plot in Figure 2(D)), 'TRIM28' (tripartite motif containing 28, a co-factor for 'CEBPB' involved in certain signal transduction pathways of glucocorticoids and inflammatory cytokines [52]) and 'CEBPB' are found to regulate the expressions of each other, and their expression patterns shown in Figure 2(A) are consistent with this prediction. 'hsa-miR-191' is identified as the primary miRNA that targets on the 'CEBPB' mRNAs, through which the expressions of a number of important genes are indirectly co-regulated. More specifically, the 'CCL5' (chemokine ligand 5) gene, which encodes the RANTES chemokine that plays an active role in leukocyte recruiting and NK cell activation and proliferation (together with IFN- $\gamma$ ) [53], is found to be the co-regulation target. The 'ISG20' gene (interferon stimulated gene) is involved in the antiviral function of interferon against RNA viruses [51], and 'ABCG1' (ATP-binding cassette subfamily G member 1) is involved in macrophage cholesterol and phospholipids transport, and is crucial for IAV replications and clearance responses of immune cells [51]. Both 'ISG20' and 'ABCG1' are also the target genes co-regulated by 'hsa-miR-191' and 'CEBPB'. Such results provide further evidence for the importance of co-regulations in immune responses to influenza virus infection.

To gain a better view of the broad picture, we perform the functional enrichment analysis on each of the 303 modules using DAVID [54]. In Table 4, the enriched GO (gene ontology) biological processes [55] and the documented pathways in KEGG/REACTOME [33,56] in 25 selected modules are listed for illustration purpose; for the complete analysis results, see Additional file 5. According to Table 4, the 'cell cycle', 'cell proliferation', 'positive regulation of programmed cell death', and 'apoptosis' are enriched, which suggests an active role of the cell-cycle related functions in response to IAV infection. In addition, a number of pathways (e.g., 'Influenza Infection', 'Jak-STAT signaling pathway', 'Wnt signaling pathway', and 'MAPK signaling pathway') are enriched in respiratory epithelial cells, with which a number of important immunological functions are

associated (e.g., 'inflammatory response', 'immune response', 'I-kappaB kinase/NF-kappaB cascade', 'cell activation during immune response', and 'leukocyte activation during immune response'). For example, the interferon  $\alpha/\beta$  receptor gene (IFNAR1) as well as the entire 'Jak-STAT signaling pathway' in Module 106 are found enriched, which is consistent with the mechanism of the antiviral state development of lung epithelial cells conferred by Type I interferons [8,57].

Figures 2(A) and (B) illustrate the activated regulatory relationships in two example modules (Modules 173 and 179) at a higher level of granularity, corresponding to the two components from the left of the first row sketched in Figure 2(D). For transcriptional regulations in Figure 3, 'CCL5' is found to be regulated by 'IRF5' (interferon regulatory factor 5) and 'NFkB2' (nuclear factor kappa-B P100) besides 'CEBPB'. The induction of 'CCL5' by 'IRF5' and the regulatory roles of 'NFkB2' and 'CEBPB' in 'CCL5' expression have been reported in previous studies [53,58,59], and are successfully identified by our method. In addition, the chemokine 'IL-8' is found to be regulated by two TFs, 'NFkB2' and 'ETV4' (ETS translocation variant 4) [34,53]. However, for such many-to-one regulatory relationships, it is unclear which TF is the major regulator and further experiments are thus needed to clarify this issue. Multiple post-transcriptional regulations can also be found in Figure 3. For example, 'hsa-miR-191' is identified to suppress gene 'REPS1' (RALBP1 associated Eps domain containing 1) in Figure 3(A), and 'hsa-miR-28-5p' to suppress gene 'CIT' (rho-interacting, serine/threonine kinase 21) in Figure 3(B). 'REPS1' is involved in cytoskeletal signaling pathway and 'CIT' encodes a serine/threonine protein kinase that functions in cell division and functions to promote efficient cytokinesis [51]. The repression of the two genes by miRNAs suggests a disorder of host factors caused by IAV infection via post-transcriptional regulations. Moreover, 'IFNGR2' (interferon gamma receptor 2) suppression by 'hsa-miR-644' and 'IFNG' (interferon gamma) suppression by 'hsa-miR-26a' are identified in Modules 107 and 287, respectively. This indicates the importance of miRNAs in innate immune responses against IAV infection by interfering IFN production and signaling [53,60]. As an example of co-regulations, the 'hsa-miR-28-5p'-'BRCA1'-'TUBB' regulation motif can be seen in Figure 3(B). 'TUBB' (tubulin beta class I) is a major constituent of microtubules [51] in the REACTOME 'Cell cycle' pathway [56], and is also a known human host proteins for IAV replication [21]. 'BRCA1' is well known as a key mediator for repairing DNA damage by maintaining the genomic integrity [51]. A better understanding of the activated regulation among 'hsa-miR-28-5p', 'BRCA1' and 'TUBB'



**Figure 2** Expression patterns of selected TFs (A), miRNAs (B) and target genes (C) in the combined co-regulation motifs (D). The first motif cluster in (D) contains the motifs 'CEBPB'-hsa-miR-191-'CCL5'/TRIM28'/ISG20'/ABCG1' etc., and the second motif cluster contains 'BRCA1'-hsa-miR-28-5p-'TUBB'/POLR2A'.



**Table 4 Functional terms of GO biological processes and KEGG/REACTOME pathways enriched in 25 selected regulatory modules (FDR < 0.05)**

Module	Term	Representative genes	p-value	FDR	
Module103	GO:0008283 ~ cell proliferation	CREBBP	0.000353	0.00498	
	GO:0006357 ~ regulation of transcription from RNA polymerase II promoter	CREBBP, EP300	0.003488	0.04824	
Module106	GO:0019221 ~ cytokine-mediated signaling pathway	IRAK4	0.001318	0.0192	
	hsa04630:Jak-STAT signaling pathway	IFNAR1, STAT2, IRF9	0.002251	0.0177	
Module130	GO:0006915 ~ apoptosis	HSPE1	0.001585	0.01762	
Module159	GO:0007249 ~ I-kappaB kinase/NF-kappaB cascade	IRF3, TRAF2, TICAM1, BCL3	3.58E-08	5.43E-07	
	GO:0043068 ~ positive regulation of programmed cell death	TRAF2, TICAM1, BCL3	1.46E-07	2.21E-06	
	GO:0002263 ~ cell activation during immune response	TICAM1, BCL3	2.23E-05	0.00034	
	GO:0002366 ~ leukocyte activation during immune response	TICAM1, BCL3	2.23E-05	0.00034	
	hsa04622:RIG-I-like receptor signaling pathway	IRF3, TRAF2, MAVS	2.27E-05	0.00021	
	GO:0001819 ~ positive regulation of cytokine production	TRAF2, MAVS, TICAM1, BCL3	0.000347	0.00526	
	GO:0051251 ~ positive regulation of lymphocyte activation	TRAF2, TICAM1	0.000433	0.00655	
	GO:0045321 ~ leukocyte activation	TICAM1, BCL3	0.00047	0.00711	
	GO:0006955 ~ immune response	MAVS, TICAM1, BCL3	0.000483	0.00731	
	GO:0009615 ~ response to virus	IRF3, MAVS, TICAM1, BCL3	0.000609	0.0092	
	REACT_6900:Signaling in Immune system	IRF3, TICAM1	0.001402	0.00804	
	Module171	hsa04310:Wnt signaling pathway	JUN, MAPK10, DVL2	0.000233	0.00246
		hsa04010:MAPK signaling pathway	JUN, MAPK10, MAPK1	0.001514	0.01594
GO:0034097 ~ response to cytokine stimulus		JUN	0.00196	0.03023	
GO:0002237 ~ response to molecule of bacterial origin		JUN, MAPK1	0.002673	0.04102	
Module173	GO:0006954 ~ inflammatory response	CEBPB, CCL5, F8, BDKRB1, ITGB2, HIF1A, TF, F3, C1RL, IL8	3.00E-05	0.00048	
	GO:0009611 ~ response to wounding	CEBPB, CCL5, F8, BDKRB1, IGFBP1, ITGB2, NRG1, HIF1A, TF, F3, C1RL, IL8	5.28E-05	0.00085	
	GO:0051240 ~ positive regulation of multicellular organismal process	CCL5, PTGS2, NRG1, HIF1A, MYLK2, TF, F3, IL27RA	0.000197	0.00316	
	GO:0006952 ~ defense response	CEBPB, CCL5, F8, BDKRB1, ITGB2, HIF1A, TF, F3, IL27RA, C1RL, IL8	0.000835	0.01331	
	GO:0032101 ~ regulation of response to external stimulus	CCL5, VEGFA, PTGS2, NT5E, F3, IL8	0.001131	0.01799	
	REACT_604:Hemostasis	STX4, SLC7A7, F8, VEGFA, ITGB2, ALB, TF, F3	0.001131	0.00976	
	GO:0002526 ~ acute inflammatory response	CEBPB, F8, TF, F3, C1RL	0.0014	0.02222	
Module174	GO:0043069 ~ negative regulation of programmed cell death	CEBPB, VEGFA, NRG1, ALB, KRT18, PPT1, PCSK6, F3	0.001937	0.03062	
	GO:0051726 ~ regulation of cell cycle	SMAD3, CDK4	1.10E-05	0.00016	
	GO:0031328 ~ positive regulation of cellular biosynthetic process	SMAD3, CDK4	0.000612	0.00894	
	GO:0006350 ~ transcription	SMAD3, ASH2L	0.000854	0.01245	
Module175	GO:0000122 ~ negative regulation of transcription from RNA polymerase II promoter	CTNBN1,	0.002252	0.02972	
Module178	GO:0051252 ~ regulation of RNA metabolic process	STAT1, UBE2I, HDAC3, PIAS2, DAXX, SP100	5.05E-05	0.00076	
	REACT_11061:Signalling by NGF	HDAC3, AKT1	0.000413	0.00293	
	GO:0007049 ~ cell cycle	DAXX, UBE2I, HDAC3, AKT1	0.0005	0.00753	
	hsa04630:Jak-STAT signaling pathway	STAT1, AKT1, PIAS2	0.004991	0.04684	

**Table 4 Functional terms of GO biological processes and KEGG/REACTOME pathways enriched in 25 selected regulatory modules (FDR < 0.05) (Continued)**

Module179	GO:0006974 ~ response to DNA damage stimulus	RAD54L, FANCI, XAB2, BCCIP, BRCA1, XRCC1, EEPD1, UPF1, RAD51, TOP2A, FANCD2, RAD54B	3.39E-08	4.91E-07
	REACT_216:DNA Repair	FANCI, XAB2, BRCA1, XRCC1, POLR2K, RAD51, POLR2A, FANCD2	1.64E-07	1.25E-06
	GO:0033554 ~ cellular response to stress	FANCI, RAD54L, XAB2, EEPD1, BRCA1, XRCC1, TOP2A, RAD54B, DHX9, BCCIP, UPF1, RAD51, FANCD2	2.71E-07	3.92E-06
	GO:0007049 ~ cell cycle	FANCI, RAD54L, KIF15, BRCA1, CHTF18, RCC1, RAD54B, CIT, BCCIP, UPF1, TUBB, RAD51, FANCD2	7.43E-06	0.00011
Module180	GO:0010604 ~ positive regulation of macromolecule metabolic process	NUP62	7.41E-05	0.0011
Module182	GO:0006303 ~ double-strand break repair via nonhomologous end joining	PRKDC	1.88E-06	2.83E-05
	GO:0045935 ~ positive regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	ILF2, RELA, NFKB1, PRKDC	0.000934	0.01395
	GO:0002562 ~ somatic diversification of immune receptors via germline recombination within a single locus	PRKDC	0.001732	0.02573
Module184	GO:0022402 ~ cell cycle process	MLH1, MAP2K6	0.001302	0.01777
Module185	GO:0007243 ~ protein kinase cascade	SRC, IKKBK, NFKBIA	0.001885	0.02759
Module186	REACT_1538:Cell Cycle Checkpoints	MCM4, MCM3, MCM5, MCM7, MCM2	9.90E-23	7.84E-22
	GO:0006260 ~ DNA replication	MCM4, MCM3, MCM5, MCM7, MCM2	1.13E-21	1.51E-20
Module189	GO:0032268 ~ regulation of cellular protein metabolic process	CBS	0.00036	0.00532
Module190	GO:0008380 ~ RNA splicing	SFPQ	2.23E-06	3.19E-05
Module200	REACT_6185:HIV Infection	NUP188, NUP205	0.000592	0.00484
Module206	hsa04350:TGF-beta signaling pathway	SMURF2, SMAD6	0.001698	0.00872
Module241	GO:0006913 ~ nucleocytoplasmic transport	NXT1, RAN	0.001111	0.01645
Module244	REACT_12472:Regulatory RNA pathways	POLR2H, POLR2I, POLR2B, POLR2D, POLR2L	5.52E-10	3.65E-09
	REACT_6143:Pausing and recovery of Tat-mediated HIV-1 elongation	POLR2H, POLR2I, POLR2B, POLR2D, POLR2L	4.94E-09	3.27E-08
	REACT_6167:Influenza Infection	POLR2H, POLR2I, POLR2B, POLR2D, POLR2L	4.39E-06	2.91E-05
Module253	GO:0000398 ~ nuclear mRNA splicing, via spliceosome	HNRNPM	0.003042	0.04485
Module267	hsa04920:Adipocytokine signaling pathway	RXRB	8.48E-06	7.22E-05
Module280	GO:0006984 ~ ER-nuclear signaling pathway	PAK1, EIF2AK3	0.001649	0.02486
Module303	hsa04144:Endocytosis	VPS28	0.000182	0.00041

'Module' refers to the module indices we gave. 'Term' refers to the enriched GO terms (e.g. GO:0008283), KEGG pathways (e.g. hsa04630), and REACTOME pathways (e.g. REACT\_6900). Certain genes in the modules are listed as 'Representative genes'. 'p-value' and 'FDR' shows the statistical significance of the results.

could be important to designing the miRNA gene therapeutic trials targeting IAV host factors [21].

## Conclusions

In this study, we proposed a novel framework for systematic identification of transcriptional and post-transcriptional regulations (SITPR) using curated knowledge, sequence-based predictions and time-course expression data. Briefly, a comprehensive background regulatory network, which is

not cell type- or disease-specific, was constructed at the genome-wide scale using information from ~20 databases and predictions from representative algorithms. After dividing the background network into smaller modules, we further proposed the constrained LASSO method and applied it with the dynamic Bayesian network model to each of the modules to determine the activated regulatory relationships based on the time-course expression data of mRNAs and miRNAs. The simulation studies clearly

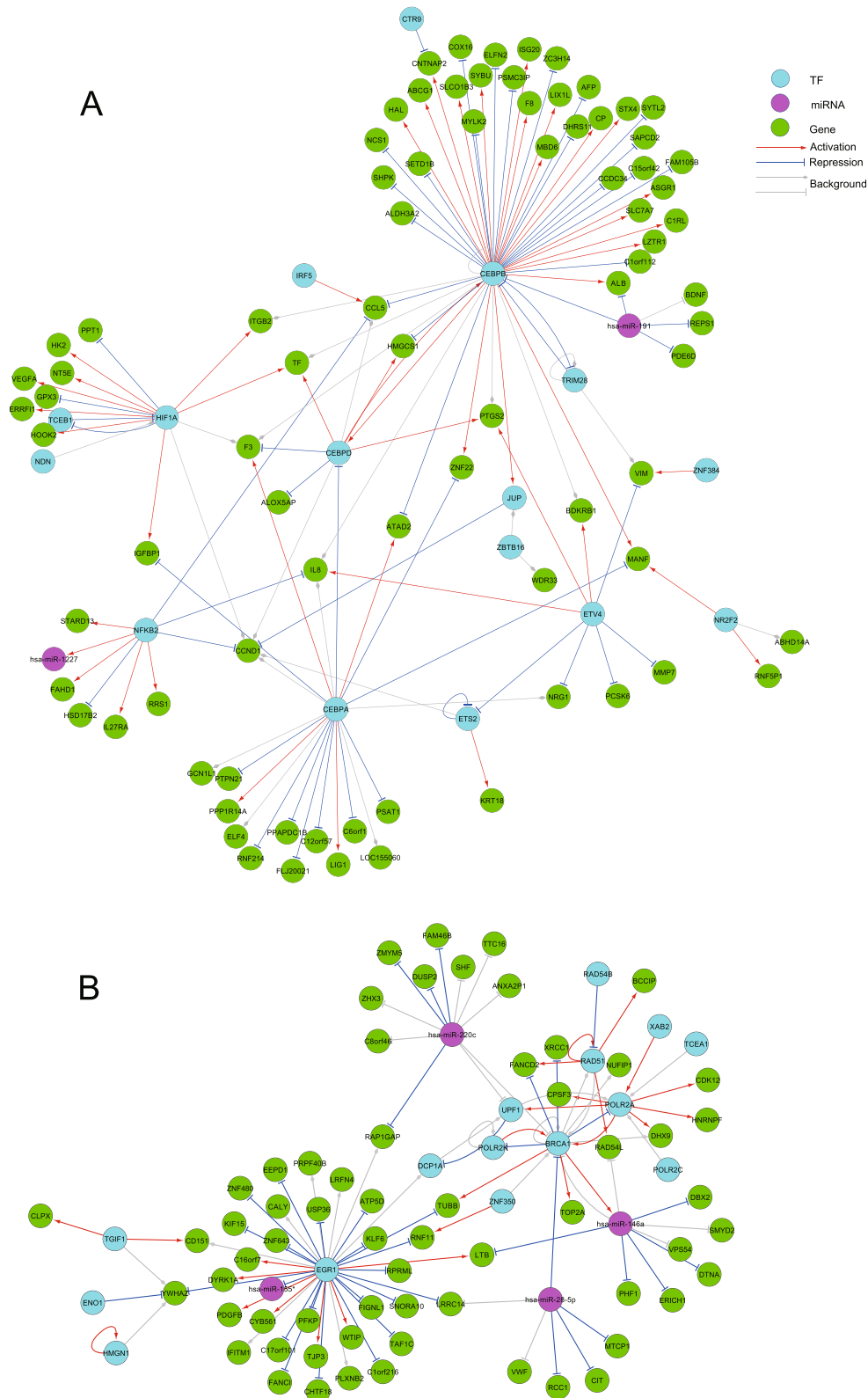


Figure 3 (See legend on next page.)

(See figure on previous page.)

**Figure 3 The activated regulatory relationships in two example modules.** TFs, miRNAs and genes are in cyan, magenta and green, respectively. The up-regulation and down-regulation are labeled in red and blue, respectively. The background regulatory relationships which are not activated are in gray. The 'arrow', 'T' and 'diamond' shapes of edge terminals denote to up-, down-, and undetermined- regulations, respectively. **(A)** Module173. This module contains the co-regulation motifs 'CEBPB'-hsa-miR-191'-CCL5' and 'CEBPB'-hsa-miR-191'-ALB'/ISG20'. It also contains some two-node regulatory motifs, e.g., 'ETS2'-ETS2', 'NFKB2'-hsa-miR-1227', 'IRF5'-CCL5' and 'hsa-miR-191'-REPS1'. **(B)** Module179. This module contains the regulatory motifs 'BRCA1'-has-miR-28-5p'-TUBB'/POLR2A', 'EGR1'-hsa-miR-155\*', 'EGR1'-hsa-miR-146a'-LTB' and 'BRCA1'-hsa-miR-146a'-PHF1'.

suggest the superiority of the proposed framework over other representative approaches in the context of identifying activated regulatory relationships from a given background network. The SITPR framework was then applied to real data to identify the activated regulatory relationships among TFs, miRNAs and target genes in human respiratory epithelial cells in response to H1N1 influenza virus infection.

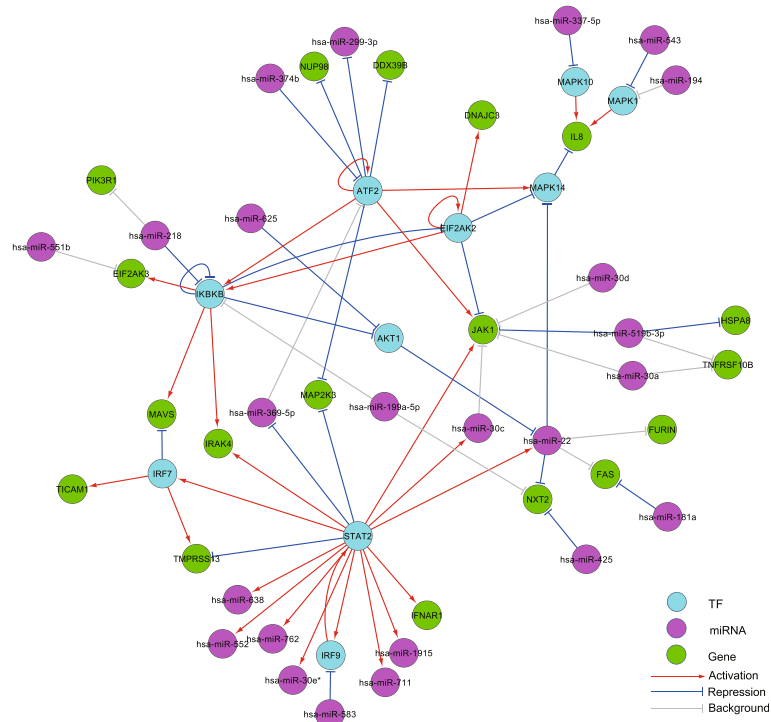
Different from many existing methods for inferring gene regulatory network that are purely data-driven [17,27,28], we incorporated curated knowledge and utilized sequence information besides time-course gene expression data. Including curated knowledge can improve the accuracy of identifying genuine regulatory relationships [30,32,61], and the incorporation of predicted potential regulatory interactions among TFs, miRNAs and genes allows to discover novel regulatory relationships between regulators and targets. Furthermore, the way we constructed the background network (e.g., using predictions of sequence binding instead of associations between expression data [42,62,63]) and the use of the constrained LASSO help to control the false positive rate in identifying activated regulatory relationships, as suggested in [64,65]. Also, since we considered the transcriptional and post-transcriptional regulations simultaneously, the identified activated regulatory relationships could be less biased [62,63,66,67]. It should be mentioned that, limited by data availability and algorithm prediction accuracy, our background network built in this study is not thorough; however, the SITPR framework is flexible and allows us to update the background network whenever new knowledge is discovered or better prediction algorithms are devised.

Using the proposed framework, the transcription regulatory landscape was depicted for human respiratory epithelial cells infected by H1N1 influenza A virus. The transcriptional and post-transcriptional regulations were simultaneously identified and analyzed for airway epithelial cells in response to IAV infection. Besides pairwise interactions, we paid particular attention to regulation motifs and their synergistic effects on gene expressions. We identified ten regulation motifs using FANDOM and selectively presented and discussed some important motifs activated during IAV infection, which could suggest potential targets for influenza treatment or vaccination. It should be addressed that the regulation motifs we

defined in this study are not only the functional building blocks but also the topological elements in a complex regulatory network. Identification of such motifs can provide the basis for advanced analyses such as differential network biology [68].

The SITPR is also applicable to a pre-defined gene set. For example, Figure 4 illustrates the activated regulatory relationships in the KEGG IAV pathway, in which we included the genes contained in this pathway as well as the neighboring miRNAs in our background network. One can tell from Figure 4 that 'STAT2' (signal transducer and activator of transcription 2) directly regulates the expressions of 9 different miRNAs, 5 target genes and two other TFs, suggesting a highly influential role of 'STAT2'. In addition, we found that 'STAT2', 'IRF9' (interferon regulatory factor 9) and 'hsa-miR-583' form a co-regulation motif. 'STAT2', 'hsa-miR-519b-3p' and 'JAK1' (Janus kinase 1) also forms another co-regulation motif. 'STAT2', 'JAK1' and 'IRF9' are known to be the key regulators in the 'Jak-STAT signaling pathway', which plays a critical role in regulating immune responses [33]. The co-regulation motif 'MAPK1'-hsa-miR-543'-IL8' was identified, which may indicate an important role of miRNA 'has-miR-543' in controlling the production of the chemokine 'IL8'. In short, one obvious benefit of analyzing a pre-defined gene set is that the network dimension is significantly reduced such that SITPR can achieve a better accuracy, as suggested by our simulation studies. Loveday et al. [14], in which the time-course expression data used in this study were generated, also conducted a regulation study on a pre-defined gene set. However, they selected hundreds of differentially expressed genes and only focused on the post-transcriptional regulations between miRNAs and target genes. Also, the regulatory relationships identified in Loveday et al. [14] were primarily based on the baseline and the data on hour 8. Therefore, although there exist a few regulatory relationships identified in both our study and Loveday's (e.g., 'hsa-miR-328' and 'PSME3'), the majority of the regulatory interactions identified in the two studies are different due to all the reasons described above.

In summary, the background regulatory network constructed in this study provides a start point for the identification of condition-specific regulatory interactions. Together with the selected dimension reduction techniques, the SITPR framework is formed to achieve a



**Figure 4 The activated regulatory network in the KEGG IAV gene set.** The genes in the IAV pathway together with their neighbors from the background network were analyzed. The importance of miRNAs can be told from the co-regulation motifs, such as 'STAT2'-'hsa-miR-583'-'IRF9', 'ATF2'-'hsa-miR-374b'-'JAK1', 'IKKβ'-'hsa-miR-218'-'AKT1', and 'MAPK1'-'hsa-miR-543'-'IL8'.

better accuracy at an affordable computing cost. Also, for human respiratory epithelial cells in response to influenza A virus infection, the results generated using SITPR reveal a number of interesting activated regulatory interactions and provide useful clues for future experimental design and validation.

## Methods

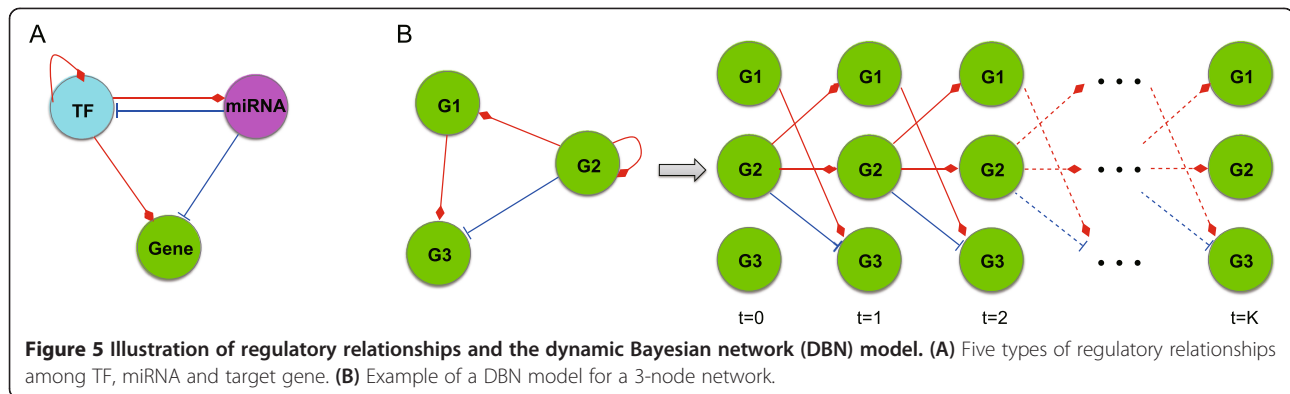
### Expression data and experiments

The time-course expression data of mRNA and miRNA in human A549 cells infected with influenza H1N1 virus (A/Mexico/InDRE4487/2009) at a multiplicity of infection (MOI) of 0.1 were downloaded from NCBI GEO [69] (GSE36553 and GSE36461). The samples of mRNA and miRNA were hybridized on Illumina HumanHT-12 v3 BeadChips and Febit miRBase 14 Geniom miRNA Biochip, respectively. Data at six time points (0, 4, 8, 24, 48, and 72 hours post infection) were collected with six replicates at each time point [14]. However, one chip sample on hour 4 was found to be degraded [14], so only 35 samples were usable for analysis. More details of the experiment and raw data preprocessing can be found in the original manuscript [14]. We mapped the probeset IDs to NCBI official gene symbols using the GEO annotation file. When multiple probesets were mapped to the same gene, the

probeset with the maximum inter quartile expression range was selected.

### Construction of the background regulatory network

TFs and miRNAs are the major transcriptional and post-transcriptional regulators, respectively [18,19]. Simultaneously considering the interplays among TFs, miRNAs and target genes could result in a more accurate identification of regulatory relationships [62]. Therefore, five types of regulatory relationships need to be accommodated in the background network as illustrated in Figure 5(A): 'TF-gene', 'miRNA-gene', 'TF-miRNA', 'miRNA-TF' and 'TF-TF' self-regulation. For this purpose, we collected a large number of documented regulatory relationships from about 20 public databases and predicted the potential regulations among TFs, miRNAs and target genes by scanning the sequence binding motifs. It should be addressed that the background regulatory network is not cell type or disease specific since it is constructed at the genome-wide scale. Also, the background network allows the discovery of novel regulations by including potential regulations. Instead of associations in gene expression data, the potential regulations are predicted by matching the sequence binding motifs of targets, which is helpful to reduce the false positive rate (FPR) in the inferred regulations [70].



Specifically, human TFs were compiled from FANTOM [71], TRANSFAC [35] and JASPAR [36]; human miRNAs were downloaded from miRBase [72]; and human genes and annotations were from NCBI GenBank and RefSeq [73]. For convenience, all the human TFs and genes were denoted using their NCBI official gene symbols. For the ‘TF-gene’ and ‘TF-TF’ regulations, we downloaded the curated regulations from TRED [34] and KEGG [33], and the regulatory DNA elements of transcriptional factor binding sites (TFBS) were extracted from TRANSFAC [35] and JASPAR [36]. For binding motifs in TRANSFAC, we identified the putative binding regions of TFs in human genome using the tables *tfbsConsSites* and *tfbsConsFactors* from UCSC Genome Browser [74], as suggested by the ENCODE project [42]. When mapping to the human genome (UCSC h19 human genome assembly), the binding regions were searched within the promoter region from 5-kb upstream to 1-kb downstream of the transcription start site (TSS) annotated by RefSeq. We used the default Z-score of 2.33 to control the type I error in binding region prediction. For JASPAR TFBS annotation, we used the *MotifFeatures* and *AnnotatedFeatures* tables from Ensembl [75] and extracted the TFs and their binding sites regions in the human genome. Again, we mapped the binding sites to human genome and identified the regulatory relationships between TFs and their potential targets. For completeness, we also extracted the TF and their target proteins from human protein-protein interaction data in HPRD [76] and KEGG [33], and mapped such interactions to the ‘TF-gene’ and ‘TF-TF’ regulations, which allows a more thorough and systematic exploration of the regulatory interactions as suggested in previous studies [62]. For the ‘TF-miRNA’ regulations, the previously-reported TF-miRNA interactions were downloaded from TransmiR [77]. We used a procedure similar to that for the ‘TF-gene’ and ‘TF-TF’ regulations to identify the regulations between TFs and miRNA’s corresponding genes. For the ‘miRNA-gene’ and ‘miRNA-TF’ regulations, the curated miRNA-gene interactions are available in public databases such as Tarbase [78], miRecords [79] and miRTarBase [80]. There

are also numerous computing methods for predicting miRNA targets, such as TargetScan [18], miRanda [81], PicTar [82], microT [83] and MicroCosm [72]. We downloaded the experimentally-confirmed miRNA-gene interactions from the three public databases, and the potential interactions were kept only if they were predicted by at least two of the five algorithms mentioned above.

Finally, we built a background regulatory network for human with 23,079 nodes and 369,277 edges, consisting of 1,456 TFs, 1,904 miRNAs and 19,719 target genes. The statistics and more details of the background network are available in Additional file 1. For the real data we used in the Results Section, 37,586 mRNAs and 904 miRNAs were measured in the experiments. The overlap between the background network and the experiment data contains 18,964 nodes (1,441 TFs, 881 miRNAs, and 16,642 target genes) and 335,963 interactions. All the data and codes can be accessed at our website <http://doc.aporc.org/wiki/SITPR>.

### Dimension reduction

Although the number of edges in the background network is significantly smaller than that of a fully connected network, the computing cost is still prohibitively high if we directly use the entire background to infer the activated regulatory relationships. We thus considered several data-based and topological feature based methods for dimension reduction.

First, we only kept the differentially expressed genes during viral infection, which were identified at a significance level of 0.05 using our functional PCA approach particularly designed for time-course data [40]. Second, we filtered out the regulatory pairs if the initial change in the regulator gene expression lags behind that in the target gene expression [38]. For this purpose, we clustered the time-course expression data of mRNAs and miRNAs using STEM [39] and then determined the time of initial change in gene expression for each cluster. That is, within each cluster, the

data at the same time point are independent replicates as suggested by the experiment design [14]; therefore, the two-sample t-test with Bonferroni correction was conducted to compare the expression levels at  $t = 1, \dots, K$  ( $K=5$ ) with that at  $t=0$ , and the first time point with a  $p$ -value less than 0.01 was deemed as the initial change time point. Obviously, the target genes as well as TFs in the same cluster will have the same time of initial change. Then if the initial change time of a TF is later than that of its potential target genes, the corresponding TF-gene interactions are filtered out [38]. Third, we filtered out the regulatory pairs that have no any association in their expression data. For this purpose, we calculated the MIC scores (and the associated  $p$ -values) [41], which can measure both linear and nonlinear associations. The associations that have a  $p$ -value less than 0.05 were kept, and the MIC scores were assigned to the corresponding regulatory interactions as the edge weights for later use.

For further dimensional reduction, the weighted background regulatory network, which was generated at the third step in the previous paragraph, was divided into multiple smaller modules using the fast community detection algorithm [37] such that the connections within modules are much denser than those between modules [84]. That is, the nodes within each module are tightly connected by many edges, while the nodes in different modules are connected by a smaller number of edges. Therefore, the generated modules can be deemed as the functional blocks of the regularity network [85,86]. We obtained 303 modules with the number of nodes in these modules ranging from 2 to ~200 and the number of edges ranging from 1 to ~300. Finally, it should be mentioned that the edge weights are not used in any further analysis.

### DBN and constrained LASSO

Since miRNAs repress the translation from mRNAs to proteins and TFs may regulate themselves, a directed graph model with cycles is necessarily needed to model each regulatory module. The dynamic Bayesian network (DBN) [38,87] was considered in this study, as illustrated in Figure 5(B). Let  $X^t = (x_1^t, \dots, x_n^t)^T$  denote the gene expression vector of  $n$  genes at time  $t$  ( $t = 0, \dots, K$ ), the joint distribution function can be given as follows [30]

$$f(X^0, \dots, X^K) = f(X^0) \prod_{t=1}^K \prod_{i=1}^n f(x_i^t | \text{parent}(x_i^t)) \quad (2)$$

where we assume that  $X^{t+1}$  only depends on  $X^t$  (the Markov chain property) and the form of dependence can be described by a linear model [38,61,87]. More specifically, we used the model

$$X^{t+1} = \mathbf{A} \cdot X^t + E \quad (3)$$

where  $\mathbf{A}$  is a matrix of constant regulatory coefficients,  $E$  is the error vector that follows a multivariate normal distribution  $N(0, \Sigma)$  with  $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$ . In matrix  $\mathbf{A}$ , let  $a_{i,j}$  denote the regulatory coefficient between the regulator  $x_j$  and the target  $x_i$ . According to this definition,  $a_{i,j} = 0$  if there exists no regulation between  $x_j$  and  $x_i$  in the background network,  $a_{i,j} > 0$  when a positive regulation exists, and  $a_{i,j} < 0$  when a negative regulation occurs.

To control the false positive rate in model selection for a high-dimensional linear regression problem as in Eq. (3), the LASSO method can be considered [64]. However, the existing LASSO formulations such as the adaptive LASSO [88] or the generalized LASSO [65] cannot directly accommodate the constraints posed on  $a_{i,j}$ ; for example, the regulatory coefficients associated with miRNAs are always negative. Therefore, we propose the so-called constrained LASSO, which is given as follows

$$\begin{aligned} \min \quad & \|\tilde{\mathbf{X}} \cdot \tilde{\mathbf{A}} - \tilde{\mathbf{B}}\|_{L_2}^2 + \lambda \|\tilde{\mathbf{A}}\|_{L_1}, \\ \text{s.t.} \quad & \tilde{\mathbf{G}} \cdot \tilde{\mathbf{A}} \geq 0, \\ & \tilde{\mathbf{S}} \cdot \tilde{\mathbf{A}} \leq 0, \\ & \tilde{\mathbf{E}} \cdot \tilde{\mathbf{A}} = 0, \end{aligned} \quad (4)$$

where  $\tilde{\mathbf{A}}_{n^2 \times 1} = (a_{1,1}, \dots, a_{1,n}, \dots, a_{n,1}, \dots, a_{n,n})^T$  denotes the coefficient vector generated by stacking the columns of  $\mathbf{A}$ ,  $\tilde{\mathbf{X}}_{nK \times n^2}$  denotes the direct sum of the matrices  $[X^0, \dots, X^0]_{n \times n}^T, \dots, [X^{K-1}, \dots, X^{K-1}]_{n \times n}^T$ , and  $\tilde{\mathbf{B}}_{nK \times 1}$  is the vector generated by stacking the expression vectors  $X^1, \dots, X^K$ . In addition,  $\tilde{\mathbf{G}}_{n^2 \times n^2}$  is a diagonal matrix with its  $(n(j-1) + i, n(j-1) + i)$  entry being 1 only if  $a_{i,j} \geq 0$ ; otherwise, the diagonal entry is set to zero.  $\tilde{\mathbf{S}}_{n^2 \times n^2}$  and  $\tilde{\mathbf{E}}_{n^2 \times n^2}$  are constructed in a similar manner, representing negative and equality constraints, respectively. Also, experimentally-confirmed regulatory relationships correspond to constraints  $a_{i,j} \neq 0$ ; for such constraints, we introduced two nonnegative variables such that  $a_{i,j} = a_{i,j}^+ - a_{i,j}^-$  and the constraints on  $a_{i,j}^+$  and  $a_{i,j}^-$  can be incorporated into  $\tilde{\mathbf{G}}$ . In addition,  $\lambda \geq 0$  is the penalty coefficient and the optimal  $\lambda$  value can be determined using a cross-validation procedure. It has been shown by previous studies that the optimal  $\lambda$  value should lie within the interval  $[0, \lambda^{\max}]$ , where  $\lambda^{\max} = \|2\tilde{\mathbf{X}}^T \tilde{\mathbf{B}}\|_{L_\infty}$  [89]. We thus search the equally-spaced grids in this range using the five-fold cross validation [64] for the optimal value, which corresponds to the minimum square prediction error [64,89]. Finally, it should be mentioned that the optimization problem in Eq. (4) is solved using quadratic programming [64,88].

## Additional files

### Additional file 1: Text S1, Figure TS1-TS5 and Table TS1-TS4

Construction of human regulatory background network.

**Additional file 2: Table S1.** Details of the simulated regulatory networks with 50 (A) and 100 (B) nodes.

**Additional file 3: Table S2.** The characterized background regulatory network (A) and the activated regulatory relationships identified by SITPR (B).

**Additional file 4: Table S3.** The full list of the activated co-regulation motifs.

**Additional file 5: Table S4.** The full list of enriched GO terms and documented pathways in the identified activated regulatory network during IAV infection.

## Competing interests

The authors declare no competing interests.

## Authors' contributions

ZPL proposed the idea, carried out the experiments, wrote the program and drafted the manuscript. HW coordinated this study and revised the manuscript. JZ contributed to the biological interpretation of the results and contributed to the manuscript writing. HM proposed the idea, oversaw the study, and significantly contributed to manuscript preparation. All authors read and approved the final manuscript.

## Acknowledgements

This work was partially supported by University of Rochester Center for Biodefense Immune Modeling grant HHSN272201000055C (NIH/NIAID), University of Rochester Center for AIDS Research grant P30AI078498 (NIH/NIAID), NIH grant R01 GM100788, and National Natural Science Foundation of China (NSFC) under Grant No. 31100949 and the Fundamental Research Funds of Shandong University under Grant No. 2014TB006.

## Author details

<sup>1</sup>Department of Biomedical Engineering, School of Control Science and Engineering, Shandong University, Jinan, Shandong 250061, China.

<sup>2</sup>Department of Biostatistics and Computational Biology, University of Rochester, Rochester, NY 14642, USA. <sup>3</sup>Department of Microbiology and Immunology, University of Rochester, Rochester, NY 14642, USA.

Received: 26 June 2014 Accepted: 23 September 2014

Published: 4 October 2014

## References

- World Health Organization (WHO): **Influenza Factsheet No. 211**. 2009. <http://www.who.int/mediacentre/factsheets/fs211/en>.
- Fraser C, Donnelly CA, Cauchemez S, Hanage WP, Van Kerkhove MD, Hollingsworth TD, Griffin J, Baggaley RF, Jenkins HE, Lyons EJ, Jombart T, Hinsley WR, Grassly NC, Balloux F, Ghani AC, Ferguson NM, Rambaut A, Pybus OG, Lopez-Gatell H, Alpuche-Aranda CM, Chapela IB, Zavala EP, Guevara DME, Checchi F, Garcia E, Hugonnet S, Roth C, WHO Rapid Pandemic Assessment Collaboration: **Pandemic potential of a strain of influenza A (H1N1): Early findings**. *Science* 2009, **324**(5934):1557–1561.
- Taniguchi T, Takaoka A: **The interferon- $\alpha/\beta$  system in antiviral responses: a multimodal machinery of gene regulation by the IRF family of transcription factors**. *Curr Opin Immunol* 2002, **14**(1):111–116.
- Alexander WS, Hilton DJ: **The role of Suppressors of Cytokine Signaling (SOCS) proteins in regulation of the immune response**. *Annu Rev Immunol* 2004, **22**(1):503–529.
- Lindsay MA: **microRNAs and the immune response**. *Trends Immunol* 2008, **29**(7):343–351.
- Basso K, Margolin AA, Stolovitzky G, Klein U, Dalla-Favera R, Califano A: **Reverse engineering of regulatory networks in human B cells**. *Nat Genet* 2005, **37**(4):382–390.
- Yosef N, Shalek AK, Gaublotte JT, Jin H, Lee Y, Awasthi A, Wu C, Karwacz K, Xiao S, Jorgolli M, Gennert D, Satija R, Shakya A, Lu DY, Trombetta JJ, Pillai MR, Ratcliffe PJ, Coleman ML, Bix M, Tantin D, Park H, Kuchroo VK, Regev A: **Dynamic regulatory network controlling TH17 cell differentiation**. *Nature* 2013, **496**(7446):461–468.
- Sanders CJ, Doherty PC, Thomas PG: **Respiratory epithelial cells in innate immunity to influenza virus infection**. *Cell Tissue Res* 2011, **343**(1):13–21.
- Geiss GK, Salvatore M, Tumpey TM, Carter VS, Wang X, Basler CF, Taubenberger JK, Bumgarner RE, Palese P, Katze MG, Garcia-Sastre A: **Cellular transcriptional profiling in influenza A virus-infected lung epithelial cells: The role of the nonstructural NS1 protein in the evasion of the host innate defense and its potential contribution to pandemic influenza**. *Proc Natl Acad Sci* 2002, **99**(16):10736–10741.
- Hayashi S, Jibiki I, Asai Y, Gon Y, Kobayashi T, Ichiwata T, Shimizu K, Hashimoto S: **Analysis of gene expression in human bronchial epithelial cells upon influenza virus infection and regulation by p38 mitogen-activated protein kinase and c-Jun-N-terminal kinase**. *Respirology* 2008, **13**(2):203–214.
- Tatebe K, Zeytun A, Ribeiro R, Hoffmann R, Harrod K, Forst C: **Response network analysis of differential gene expression in human epithelial lung cells during avian influenza infections**. *BMC Bioinformatics* 2010, **11**(1):1–15.
- Chakrabarti A, Vipat V, Mukherjee S, Singh R, Pawar S, Mishra A: **Host gene expression profiling in influenza A virus-infected lung epithelial (A549) cells: a comparative analysis between highly pathogenic and modified H5N1 viruses**. *Virology* 2010, **7**(1):219.
- Shapira SD, Gat-Viks I, Shum BOV, Dricot A, de Grace MM, Wu L, Gupta PB, Hao T, Silver SJ, Root DE, Hill DE, Regev A, Hacohen N: **A physical and regulatory map of host-influenza interactions reveals pathways in H1N1 infection**. *Cell* 2009, **139**(7):1255–1267.
- Loveday EK, Svinti V, Diederich S, Pasick J, Jean F: **Temporal- and strain-specific host microRNA molecular signatures associated with swine-origin H1N1 and avian-origin H7N7 influenza A virus infection**. *J Virol* 2012, **86**(11):6109–6122.
- Tsang J, Zhu J, van Oudenaarden A: **MicroRNA-mediated feedback and feedforward loops are recurrent network motifs in mammals**. *Mol Cell* 2007, **26**(5):753–767.
- Martinez NJ, Ow MC, Barrasa MI, Hammell M, Sequerra R, Doucette-Stamm L, Roth FP, Ambros VR, Walhout AJ: **A C. elegans genome-scale microRNA network contains composite feedback motifs with high flux capacity**. *Genes Dev* 2008, **22**(18):2535–2549.
- Marbach D, Costello JC, Kuffner R, Vega NM, Prill RJ, Camacho DM, Allison KR, Consortium D, Kellis M, Collins JJ, Stolovitzky G: **Wisdom of crowds for robust gene network inference**. *Nat Methods* 2012, **9**(8):796–804.
- Bartel DP: **MicroRNAs: genomics, biogenesis, mechanism, and function**. *Cell* 2004, **116**(2):281–297.
- He L, Hannon GJ: **MicroRNAs: small RNAs with a big role in gene regulation**. *Nat Rev Genet* 2004, **5**(7):522–531.
- Josset L, Belsler JA, Pantin-Jackwood MJ, Chang JH, Chang ST, Belisle SE, Tumpey TM, Katze MG: **Implication of inflammatory macrophages, nuclear receptors, and interferon regulatory factors in increased virulence of pandemic 2009 H1N1 influenza A virus after host adaptation**. *J Virol* 2012, **86**(13):7192–7206.
- Konig R, Stertz S, Zhou Y, Inoue A, Hoffmann HH, Bhattacharyya S, Alamares JG, Tscherne DM, Ortigoza MB, Liang YH, Gao QS, Andrews SE, Bandyopadhyay S, De Jesus P, Tu BP, Pache L, Shih C, Orth A, Bonamy G, Miraglia L, Ideker T, Garcia-Sastre A, Young JAT, Palese P, Shaw ML, Chanda SK: **Human host factors required for influenza virus replication**. *Nature* 2010, **463**(7282):813–817.
- Butte AJ, Kohane IS: **Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements**. *Pac Symp Biocomput* 2000, **5**:418–429.
- Friedman N, Linial M, Nachman I, Pe'er D: **Using Bayesian networks to analyze expression data**. *J Comput Biol* 2000, **7**(3–4):601–620.
- Yeung MK, Tegner J, Collins JJ: **Reverse engineering gene networks using singular value decomposition and robust regression**. *Proc Natl Acad Sci U S A* 2002, **99**(9):6163–6168.
- Zhang B, Horvath S: **A general framework for weighted gene co-expression network analysis**. *Stat Appl Genet Mol Biol* 2005, **4**:17.
- Wu H, Lu T, Xue H, Liang H: **Sparse additive ordinary differential equations for gene regulatory network modeling**. *J Am Stat Assoc* 2014, **109**(506):700–716.
- Bansal M, Belcastro V, Ambesi-Impiombato A, di Bernardo D: **How to infer gene networks from expression profiles**. *Mol Syst Biol* 2007, **3**:78.



28. Marbach D, Prill RJ, Schaffter T, Mattiussi C, Floreano D, Stolovitzky G: **Revealing strengths and weaknesses of methods for gene network inference.** *Proc Natl Acad Sci U S A* 2010, **107**(14):6286–6291.
29. Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM: **A census of human transcription factors: function, expression and evolution.** *Nat Rev Genet* 2009, **10**(4):252–263.
30. Saito S, Aburatani S, Horimoto K: **Network evaluation from the consistency of the graph structure with the measured data.** *BMC Syst Biol* 2008, **2**:84.
31. Ideker T, Dutkowski J, Hood L: **Boosting signal-to-noise in complex biology: prior knowledge is power.** *Cell* 2011, **144**(6):860–863.
32. Mukherjee S, Speed TP: **Network inference using informative priors.** *Proc Natl Acad Sci U S A* 2008, **105**(38):14313–14318.
33. Kanehisa M, Goto S: **KEGG: kyoto encyclopedia of genes and genomes.** *Nucleic Acids Res* 2008, **28**(1):27–30.
34. Zhao F, Xuan Z, Liu L, Zhang MQ: **TRED: a Transcriptional Regulatory Element Database and a platform for in silico gene regulation studies.** *Nucleic Acids Res* 2005, **33**(Database issue):D103–107.
35. Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, Reuter I, Chekmenov D, Krull M, Hornischer K, Voss N, Stegmaier P, Lewicki-Potapov B, Saxel H, Kel AE, Wingender E: **TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes.** *Nucleic Acids Res* 2006, **34**(Database issue):D108–110.
36. Bryne JC, Valen E, Tang MH, Marstrand T, Winther O, da Piedade I, Krogh A, Lenhard B, Sandelin A: **JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update.** *Nucleic Acids Res* 2008, **36**(Database issue):D102–106.
37. Newman MEJ: **Fast algorithm for detecting community structure in networks.** *Phys Rev E* 2004, **69**(6):066133.
38. Zou M, Conzen SD: **A new dynamic Bayesian network (DBN) approach for identifying gene regulatory networks from time course microarray data.** *Bioinformatics* 2005, **21**(1):71–79.
39. Ernst J, Bar-Joseph Z: **STEM: a tool for the analysis of short time series gene expression data.** *BMC Bioinformatics* 2006, **7**:191.
40. Wu S, Wu H: **More powerful significant testing for time course gene expression data using functional principal component analysis approaches.** *BMC Bioinformatics* 2013, **14**:6.
41. Reshef DN, Reshef YA, Finucane HK, Grossman SR, McVean G, Turnbaugh PJ, Lander ES, Mitzenmacher M, Sabeti PC: **Detecting novel associations in large data sets.** *Science* 2011, **334**(6062):1518–1524.
42. Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan KK, Cheng C, Mu XJ, Khurana E, Rozowsky J, Alexander R, Min R, Alves P, Abyzov A, Addleman N, Bhardwaj N, Boyle AP, Cayting P, Charos A, Chen DZ, Cheng Y, Clarke D, Eastman C, Euskirchen G, Frieze S, Fu Y, Gertz J, Grubert F, Harman A, Jain P, Kasowski M, et al: **Architecture of the human regulatory network derived from ENCODE data.** *Nature* 2012, **489**(7414):91–100.
43. Newman MEJ: **The structure and function of complex networks.** *SIAM Rev* 2003, **45**(2):167–256.
44. Barabasi AL, Albert R: **Emergence of scaling in random networks.** *Science* 1999, **286**(5439):509–512.
45. Faith JJ, Hayete B, Thaden JT, Mogno I, Wierzbowski J, Cottarel G, Kasif S, Collins JJ, Gardner TS: **Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles.** *PLoS Biol* 2007, **5**(1):e8.
46. Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Dalla Favera R, Califano A: **ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context.** *BMC Bioinformatics* 2006, **7**(Suppl 1):S7.
47. Huynh-Thu VA, Irtuthum A, Wehenkel L, Geurts P: **Inferring regulatory networks from expression data using tree-based methods.** *PLoS One* 2010, **5**(9):e12776.
48. Haury AC, Mordelet F, Vera-Licona P, Vert JP: **TIGRESS: Trustful Inference of Gene REgulation using Stability Selection.** *BMC Syst Biol* 2012, **6**:145.
49. Shen-Orr SS, Milo R, Mangan S, Alon U: **Network motifs in the transcriptional regulation network of Escherichia coli.** *Nat Genet* 2002, **31**(1):64–68.
50. Wernicke S, Rasche F: **FANMOD: a tool for fast network motif detection.** *Bioinformatics* 2006, **22**(9):1152–1153.
51. Rebhan M, Chalifa-Caspi V, Prilusky J, Lancet D: **GeneCards: a novel functional genomics compendium with automated data mining and query reformulation support.** *Bioinformatics* 1998, **14**(8):656–664.
52. Chang CJ, Chen YL, Lee SC: **Coactivator TIF1beta interacts with transcription factor C/EBPbeta and glucocorticoid receptor to induce alpha1-acid glycoprotein gene expression.** *Mol Cell Biol* 1998, **18**(10):5880–5887.
53. Julkunen I, Sareneva T, Pirhonen J, Ronni T, Melen K, Matikainen S: **Molecular pathogenesis of influenza A virus infection and virus-induced regulation of cytokine gene expression.** *Cytokine Growth Factor Rev* 2001, **12**(2–3):171–180.
54. da Huang W, Sherman BT, Lempicki RA: **Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.** *Nat Protoc* 2009, **4**(1):44–57.
55. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.** *Nat Genet* 2000, **25**(11):25–29.
56. Joshi-Tope G, Gillespie M, Vastrik I, D'Eustachio P, Schmidt E, de Bono B, Jassal B, Gopinath GR, Wu GR, Matthews L, Lewis S, Birney E, Stein L: **Reactome: a knowledgebase of biological pathways.** *Nucleic Acids Res* 2005, **33**(Database issue):D428–432.
57. Zhou Z, Hamming OJ, Ank N, Paludan SR, Nielsen AL, Hartmann R: **Type III interferon (IFN) induces a type I IFN-like response in a restricted subset of cells through signaling pathways involving both the Jak-STAT pathway and the mitogen-activated protein kinases.** *J Virol* 2007, **81**(14):7749–7758.
58. Barnes BJ, Kellum MJ, Field AE, Pitha PM: **Multiple regulatory domains of IRF-5 control activation, cellular localization, and induction of chemokines that mediate recruitment of T lymphocytes.** *Mol Cell Biol* 2002, **22**(16):5721–5740.
59. Wickremasinghe MI, Thomas LH, O'Kane CM, Uddin J, Friedland JS: **Transcriptional mechanisms regulating alveolar epithelial cell-specific CCL5 secretion in pulmonary tuberculosis.** *J Biol Chem* 2004, **279**(26):27199–27210.
60. O'Connell RM, Rao DS, Chaudhuri AA, Baltimore D: **Physiological and pathological roles for microRNAs in the immune system.** *Nat Rev Immunol* 2010, **10**(2):111–122.
61. Kim SY, Imoto S, Miyano S: **Inferring gene networks from time series microarray data using dynamic Bayesian networks.** *Brief Bioinform* 2003, **4**(3):228–235.
62. Cheng C, Yan KK, Hwang W, Qian J, Bhardwaj N, Rozowsky J, Lu ZJ, Niu W, Alves P, Kato M, Snyder M, Gerstein M: **Construction and analysis of an integrated regulatory network derived from high-throughput sequencing data.** *PLoS Comput Biol* 2011, **7**(11):e1002190.
63. Yan Z, Shah PK, Amin SB, Samur MK, Huang N, Wang X, Misra V, Ji H, Gabuzda D, Li C: **Integrative analysis of gene and miRNA expression profiles with transcription factor-miRNA feed-forward loops identifies regulators in human cancers.** *Nucleic Acids Res* 2012, **40**(17):e135.
64. Tibshirani R: **Regression shrinkage and selection via the Lasso.** *J Roy Stat Soc B Met* 1996, **58**(1):267–288.
65. Tibshirani RJ, Taylor J: **The solution path of the generalized LASSO.** *Annal Stat* 2011, **39**(3):1335–1371.
66. Setty M, Helmy K, Khan AA, Silber J, Arvey A, Neezen F, Agius P, Huse JT, Holland EC, Leslie CS: **Inferring transcriptional and microRNA-mediated regulatory programs in glioblastoma.** *Mol Syst Biol* 2012, **8**:605.
67. Guo AY, Sun J, Jia P, Zhao Z: **A novel microRNA and transcription factor mediated regulatory network in schizophrenia.** *BMC Syst Biol* 2010, **4**:10.
68. Ideker T, Krogan NJ: **Differential network biology.** *Mol Syst Biol* 2012, **8**:565.
69. Edgar R, Domrachev M, Lash AE: **Gene Expression Omnibus: NCBI gene expression and hybridization array data repository.** *Nucleic Acids Res* 2002, **30**(1):207–210.
70. Pilpel Y, Sudarsanam P, Church GM: **Identifying regulatory networks by combinatorial analysis of promoter elements.** *Nat Genet* 2001, **29**(2):153–159.
71. Ravasi T, Suzuki H, Cannistraci CV, Katayama S, Bajic VB, Tan K, Akalin A, Schmeier S, Kanamori-Katayama M, Bertin N, Carninci P, Daub CO, Forrest AR, Gough J, Grimmond S, Han JH, Hashimoto T, Hide W, Hofmann O, Kamburov A, Kaur M, Kawaji H, Kubosaki A, Lassmann T, van Nimwegen E, MacPherson CR, Ogawa C, Radovanovic A, Schwartz A, Teasdale RD, et al: **An atlas of combinatorial transcriptional regulation in mouse and man.** *Cell* 2010, **140**(5):744–752.

72. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ: **miRBase: tools for microRNA genomics.** *Nucleic Acids Res* 2008, **36**(Database issue):D154–158.
73. Benson DA, Karsch-Mizrachi I, Clark K, Lipman DJ, Ostell J, Sayers EW: **GenBank.** *Nucleic Acids Res* 2012, **40**(Database issue):D48–53.
74. Fujita PA, Rhead B, Zweig AS, Hinrichs AS, Karolchik D, Cline MS, Goldman M, Barber GP, Clawson H, Coelho A, Diekhans M, Dreszer TR, Giardine BM, Harte RA, Hillman-Jackson J, Hsu F, Kirkup V, Kuhn RM, Learned K, Li CH, Meyer LR, Pohl A, Raney BJ, Rosenbloom KR, Smith KE, Haussler D, Kent WJ: **The UCSC Genome Browser database: update 2011.** *Nucleic Acids Res* 2011, **39**(Database issue):D876–882.
75. Flicek P, Amode MR, Barrell D, Beal K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fairley S, Fitzgerald S, Gil L, Gordon L, Hendrix M, Hourlier T, Johnson N, Kahari AK, Keefe D, Keenan S, Kinsella R, Komorowska M, Koscielny G, Kulesha E, Larsson P, Longden I, McLaren W, Muffato M, Overduin B, Pignatelli M, Pritchard B, Riat HS, et al: **Ensembl 2012.** *Nucleic Acids Res* 2012, **40**(Database issue):D84–90.
76. Mishra GR, Suresh M, Kumaran K, Kannabiran N, Suresh S, Bala P, Shivakumar K, Anuradha N, Reddy R, Raghavan TM, Menon S, Hanumanthu G, Gupta M, Upendran S, Gupta S, Mahesh M, Jacob B, Mathew P, Chatterjee P, Arun KS, Sharma S, Chandrika KN, Deshpande N, Palvankar K, Raghavath R, Krishnakanth R, Karathia H, Rekha B, Nayak R, Vishnupriya G, et al: **Human protein reference database–2006 update.** *Nucleic Acids Res* 2006, **34**(Database issue):D411–414.
77. Wang J, Lu M, Qiu C, Cui Q: **TransmiR: a transcription factor-microRNA regulation database.** *Nucleic Acids Res* 2010, **38**(Database issue):D119–122.
78. Sethupathy P, Corda B, Hatzigeorgiou AG: **TarBase: A comprehensive database of experimentally supported animal microRNA targets.** *RNA* 2006, **12**(2):192–197.
79. Xiao F, Zuo Z, Cai G, Kang S, Gao X, Li T: **miRecords: an integrated resource for microRNA-target interactions.** *Nucleic Acids Res* 2009, **37**(Database issue):D105–110.
80. Hsu SD, Lin FM, Wu WY, Liang C, Huang WC, Chan WL, Tsai WT, Chen GZ, Lee CJ, Chiu CM, Chien CH, Wu MC, Huang CY, Tsou AP, Huang HD: **miRTarBase: a database curates experimentally validated microRNA-target interactions.** *Nucleic Acids Res* 2011, **39**(Database issue):D163–169.
81. John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS: **Human MicroRNA targets.** *PLoS Biol* 2004, **2**(11):e363.
82. Krek A, Grun D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, MacMenamin P, da Piedade I, Gunsalus KC, Stoffel M, Rajewsky N: **Combinatorial microRNA target predictions.** *Nat Genet* 2005, **37**(5):495–500.
83. Maragkakis M, Reczko M, Simossis VA, Alexiou P, Papadopoulos GL, Dalamagas T, Giannopoulos G, Goumas G, Koukis E, Kourtis K, Vergoulis T, Koziris N, Sellis T, Tsanakas P, Hatzigeorgiou AG: **DIANA-microT web server: elucidating microRNA functions through target prediction.** *Nucleic Acids Res* 2009, **37**(Web Server issue):W273–276.
84. Girvan M, Newman MEJ: **Community structure in social and biological networks.** *Proc Natl Acad Sci* 2002, **99**(12):7821–7826.
85. Segal E, Shapira M, Regev A, Pe'er D, Botstein D, Koller D, Friedman N: **Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data.** *Nat Genet* 2003, **34**(2):166–176.
86. Bar-Joseph Z, Gerber GK, Lee TI, Rinaldi NJ, Yoo JY, Robert F, Gordon DB, Fraenkel E, Jaakkola TS, Young RA, Gifford DK: **Computational discovery of gene modules and regulatory networks.** *Nat Biotechnol* 2003, **21**(11):1337–1342.
87. Perrin BE, Ralaivola L, Mazurie A, Bottani S, Mallet J, D'Alche-Buc F: **Gene networks inference using dynamic Bayesian networks.** *Bioinformatics* 2003, **19**(Suppl 2):ii138–148.
88. Zou H: **The adaptive Lasso and its oracle properties.** *J Am Stat Assoc* 2006, **101**(476):1418–1429.
89. Kim SJ, Koh K, Lustig M, Boyd S, Gorinevsky D: **An interior-point method for large-scale l(1)-regularized least squares.** *IEEE J Sel Top Signal Process* 2007, **1**(4):606–617.

doi:10.1186/1471-2105-15-336

**Cite this article as:** Liu et al.: Systematic identification of transcriptional and post-transcriptional regulations in human respiratory epithelial cells during influenza A virus infection. *BMC Bioinformatics* 2014 **15**:336.

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

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

Submit your manuscript at  
www.biomedcentral.com/submit

