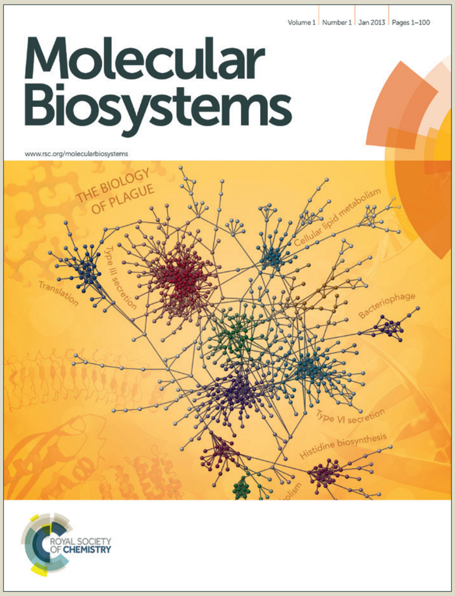

Volume 1 | Number 1 | Jan 2013 | Pages 1–100
Molecular Biosystems
www.rsc.org/molecularbiosystems
THE BIOLOGY OF PLAGUE
Cellular lipid metabolism
Type III secretion
Translation
Bacteriophage
Type VI secretion
Histidine biosynthesis
ROYAL SOCIETY OF CHEMISTRY

Molecular
BioSystems

PAPER

Exploring the associations of host genes for viral infection revealed by genome-wide RNAi and virus-host protein interactions

Dafei Xie,‡ [a] Lu Han,‡ [a] Yifu Luo,[b] Yang Liu,[a] Song He,[a] Hui Bai,*[ac] Shengqi Wang*[a] and Xiaochen Bo*[a]

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/molecularbiosystems

Genome-wide RNA interference screens have greatly facilitated identification of essential host factors (EHFs) for viral infections, whose knockdown effects significantly influence virus replication but not host cell viability. However, little has been done to link EHFs with another important host factor type, i.e., virus targeting proteins (VTPs) that viruses directly interact with for intracellular survival, hampering an integrative understanding of virus-host interactions. Using EHFs and VTPs for human immunodeficiency virus type 1 (HIV-1) and influenza A virus (IAV) infections, we found in general that despite limited overlap, EHFs and VTPs are both among the most differentially dysregulated genes in host transcriptional response to HIV and IAV infection, and notably they show consistency in regulation orientation. In human protein-protein interaction network, EHFs and VTPs both hold topologically important positions at the global center, and importantly their direct interactions are statistically significant. We also identified BRCA1 and TP53 (or SMAD3 and PIK3R1) being the most extensive VTP-interacting EHFs (or EHF-interacting VTPs) for HIV-1 and IAV, which hold great potential in deciphering specific infection features and discovery of host directed antivirals. Further, most EHFs are the upstream regulators of VTPs when mapped in the same signaling pathways, some of which present intensive cross links. Collectively, these results provide insights into functional associations of identified host gene factors for viral infections, and highlight the regulating significance of EHFs, and the necessity of their selective exploitation in confrontation to viral infections.

Introduction

Essential host factors (EHFs) are a class of host cellular genes identified from genome-wide RNA interference (RNAi) screens, the knockdown effects of which greatly influence the infection of a specific pathogen, but not the viability of host cells[1]. As a powerful high-throughput screening tool with increasingly sophisticated design, recent years have witnessed genome-wide RNAi screen generating amounting EHF data (designated as "confirmed hits" in the workflow of genome-wide RNAi screen, Fig. 1) for a dozen important human viral pathogens[2,3]. However, the growing need for in-depth biochemical and biological characterization of this newly described dataset of EHFs, and more specifically, the association between EHFs and other high-throughput screen results of virus-host interactions, has hardly been met.

With the availability of specialized host-pathogen interaction databases such as VirHostNet[4], Host-Pathogen Interaction Database[5], and human immunodeficiency virus type 1 (HIV-1) human protein interaction database[6], virus targeting proteins (VTPs) that viruses directly interact with for intracellular survival, have been identified and curated. Thus, a multi-aspect meta-analysis can highlight the associations for

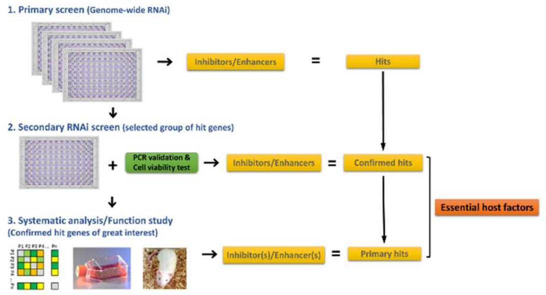

1. Primary screen (Genome-wide RNAi)
Inhibitors/Enhancers
=
Hits
2. Secondary RNAi screen (selected group of hit genes)
+
PCR validation & Cell viability test
Inhibitors/Enhancers
=
Confirmed hits
Essential host factors
3. Systematic analysis/Function study (Confirmed hit genes of great interest)
P1 P2 P3 P4 ... Pn
Inhibitor(s)/Enhancer(s)
=
Primary hits

Fig. 1 The workflow of genome-wide RNA interference (RNAi) screen to identify essential host factors (EHFs) for viral infection. The confirmed hits are validated host genes after secondary or tertiary RNAi screens whose silence causes designated phenotypes with a score above a threshold value. Generally, host genes whose silence causes the most intensive phenotype with top scores are defined as primary hits. Note: the photo of the mouse was taken in our lab by one of the co-authors.

[a] *Department of Biotechnology, Beijing Institute of Radiation Medicine, Beijing, China. Email: boxc@bmi.ac.cn, huibai13@hotmail.com, sqwang@bmi.ac.cn; Tel: +86-10-66931207*
[b] *National University of Defense Technology, Changsha, China*
[c] *No. 451 Hospital of Chinese People's Liberation Army, Xi'an, China*

† Footnotes relating to the title and/or authors should appear here.
Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

‡ These authors contributed equally to this work.
* Correspondence authors.

identified host factors of major relevance, i.e., EHFs and VTPs. It can also be expected that this may provide insights into the landscape of intricate and extensive viral dependencies on the host machinery.

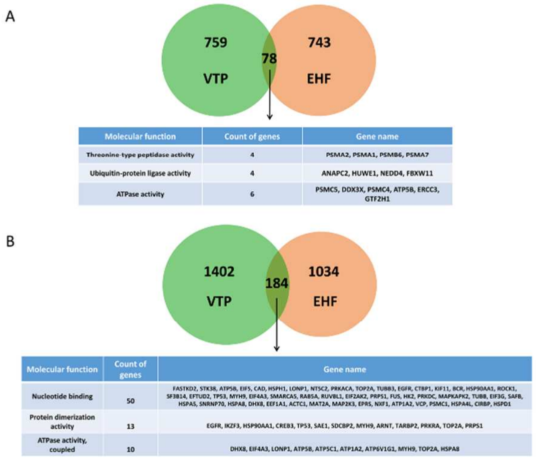
Herein, we selected two most troublesome viral pathogens, HIV-1 and influenza A virus (IAV), the EHF gene sets of which has been identified from genome-wide RNAi screens reported by different laboratories and computationally combined[7]. Meanwhile, their VTP data and host cellular gene expression profile data upon infection are publically available and sufficiently validated at different experimental settings. Through meta-analysis, the systematic associations of EHFs with VTPs were discovered at multiple aspects. This provided a comprehensive overview of EHFs of regulatory significance, and thereby aided in hypothesizing interactions among diverse host factor types for viral infection.

Results

Limited overlap between the EHF and VTP datasets

EHFs and VTPs have been acknowledged as the two main data resources to explore virus-host interactions experimentally validated by high-throughput screens[2]. Through literature investigation and data mining, we obtained a total of 821 and 1218 EHFs, as well as 837 and 1586 VTPs for HIV-1 and IAV, respectively (detailed information is provided in Table S1).

Intriguingly, we found that the overlap between datasets of EHF and VTP encoding genes for the two viruses were quite limited, i.e., 78 for HIV-1 and 184 for IAV (Fig. 2). In proportion, the overlapped genes took only 9.50% (78/821) and 9.32% (78/837) of EHFs and VTPs for HIV-1, respectively. Similarly, the overlapped genes took 15.11% (184/1218) and 11.60% (184/1586) of EHFs and VTPs for IAV, respectively. In addition, the most enriched categories of molecular function for shared genes of HIV-1 include threonine-type peptidase activity, ubiquitin-protein ligase activity and ATPase activity, while those for IAV are nucleotide binding, protein dimerization activity and coupled ATPase activity (Fig. 2). These results highly indicate the necessity to analyse the associations between EHFs and VTPs in a systematic way.



Molecular function	Count of genes	Gene name
Threonine-type peptidase activity	4	PSMA2, PSMA1, PSMB6, PSMA7
Ubiquitin-protein ligase activity	4	ANAPC2, HUWE1, NEDD4, FBXW11
ATPase activity	6	PSMC5, DDX3X, PSMC4, ATP5B, ERCC3, GTF2H1

Molecular function	Count of genes	Gene name
Nucleotide binding	50	FASTKD2, STK38, ATP5B, EIF5, CAD, HSPH1, LONP1, NT5C2, PRKACA, TOP2A, TUBB3, EGFR, CTBP1, KIF11, BCR, HSP90AA1, ROCK1, SF3B14, EFTUD2, TP53, MYH9, EIF4A3, SMARCA5, RAB5A, RUVBL1, EIF2AK2, PRPS1, FUS, HK2, PRKDC, MAPKAPK2, TUBB, EIF3G, SAFB, HSPA5, SNRNP70, HSPA8, DHX8, EEF1A1, ACTC1, MAT2A, MAP2K3, EPRS, NXF1, ATP1A2, VCP, PSMC1, HSPA4L, CIRBP, HSPD1
Protein dimerization activity	13	EGFR, IKZF3, HSP90AA1, CREB3, TP53, SAE1, SDCBP2, MYH9, ARNT, TARBP2, PRKRA, TOP2A, PRPS1
ATPase activity, coupled	10	DHX8, EIF4A3, LONP1, ATP5B, ATP5C1, ATP1A2, ATP6V1G1, MYH9, TOP2A, HSPA8

Fig. 2 The overlap between datasets of essential host factors (EHFs) and virus targeting proteins (VTPs) for (A) human immunodeficiency virus type 1 (HIV-1) and (B) influenza A virus (IAV). The most enriched gene ontology categories (molecular function, *P*<0.01) annotated for the shared genes of EHFs and VTPs for the two viruses are listed below.

EHFs and VTPs in host transcriptional response (HTR) to viral infection

After thorough screening, we designated two datasets, i.e., GSE9927 [8] and GSE31470 [9], from Gene Expression Omnibus (GEO) database reporting the genome-scale host cellular gene expression changes to represent the host transcriptional responses (HTRs) to HIV-1 and IAV infections, respectively. Specifically, GSE9927 provided the transcriptional programs of *in vivo* activated CD4(+) T cell samples from untreated HIV(+) individuals and HIV(-) individuals, and GSE31470 provided those of lung epithelial cell samples infected with low pathogenic avian influenza viruses A/WSN/33 (H1N1).

The positive and negative normalized enrichment scores (NES) indicate up- and down- regulatory consistency of subject gene set when compared with standard genome-scale host cellular gene expression changes. Taken P < 0.05 or FDR (false discovery rate) < 25% as cutoff, we found that EHF and VTP genes are both among the most significantly differentially dysregulated gene sets in the corresponding HTR to HIV and IAV infections, respectively (Table 1).To be noted, the regulation orientations of EHFs and VTPs for HIV-1 presented the opposite trend, compared to that of EHFs and VTPs for IAV. As shown in Fig. 3, the majority of EHFs and VTPs for HIV-1 were both significantly up-regulated during infection, whereas during IAV infection, the majority of EHFs and VTPs were suppressed.

To further explore the associations between EHFs and VTPs as functional regulators/coordinators, we collected the Leading-Edge Subsets(LESs)(Table S2), i.e. the core genes contribute to the ES identified by Gene set enrichment analysis (GSEA), and analysed the enriched Gene Ontology Bioprocesses(GO BPs) using DAVID online functional annotation chart tools. To ensure the reliability of analysis results, we used strict parameters (Threshold count = 10, Threshold EASE = 1.0×10-4), and we found that despite the limited overlap between EHFs and VTPs for either virus, most of their LESs enriched GO BPs showed significant overlap.

Table 1 The Gene Set Enrichment Analysis result of essential host factors (EHFs) and virus targeting proteins (VTPs) for human immunodeficiency virus type 1 (HIV-1) and influenza A virus (IAV) as compared with corresponding host cellular gene expression profiles upon infection as background.

Gene expression profile dataset	Gene Set	Size	NES	P value	FDR q-value
HIV-1 (GSE9927)	VTP	725	1.629	0.037	0.029
	EHF	693	1.327	0.017	0.085
IAV (GSE31470)	VTP	1100	-1.767	0.000	0.086
	EHF	1341	-1.609	0.000	0.086

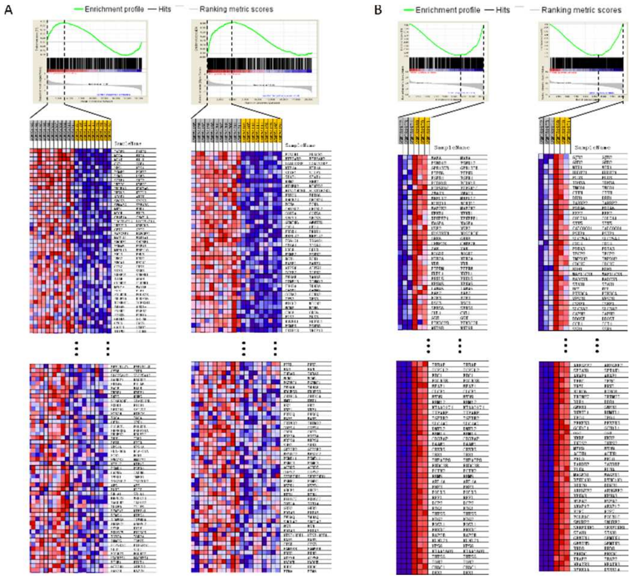


Fig. 3 Gene set enrichment analysis (GSEA) reports of enrichment of datasets of essential host factors (EHFs) and virus targeting proteins (VTPs) for human immunodeficiency virus type 1 (HIV-1) and influenza A virus (IAV) in phenotype of infection and control. Phenotypes are determined as in dataset GSE9927 and GSE31470 that respectively records uninfected and infected genome-scale host cellular gene expression changes for HIV-1 and IAV. *P* < 0.05 or FDR (false discovery rate) < 25% is taken as cutoff. Snapshot of enrichment results and heat map representation of expression changes of genes with core enrichment are shown for gene sets of (a) HIV's EHF (left) and VTP (right), and (b) IAV's EHF (left) and VTP (right). Profile of the running Enrichment Score (ES) and Positions of Gene Set Members on the Rank Ordered List are shown in the Enrichment plots. Lanes in the heat map are numbered samples, with yellow background representing control group and grey background representing infected group. Squares in the heat map are differentially colored according to the extent of gene expression changes, with red and blue representing the most up- and down-regulated genes. Note: For better illustration, only part of the genes with core enrichment in these gene sets are shown in the heat map (Detailed information are provided in Table S2).

For HIV-1, 34 and 59 GO BPs were enriched in the LESs of HIV-1's EHFs and VTPs, respectively (Table S3), and notably, 34 GO BPs were overlapped, demonstrating the intensive regulating relations of EHFs and VTPs as in the same functional categories. For example, the upregulated EHFs and VTPs of HIV-1 both showed enrichment in the functional categories of regulation of protein ubiquitination, which as reported, is crucial to HIV's escape of host clearance[10,11]. However, 60 and 71 GO BPs were enriched in the LES of IAV's EHFs and VTPs respectively, while only 5 of them were overlapped (Table S3). For example, the downregulated EHFs and VTPs of IAV's were both enriched in cell cycle, protein localization and RNA biosynthetic process, whereas many downregulated EHFs were also enriched in specific antiviral responses BPs, e.g., apoptosis, NF-kappaB cascade and MAPKKK cascade, the activation of which in early response to infection could be blocked by IAV for its intracellular survival[12]. Nonetheless, the VTP enriched bioprocesses of IAV, i.e., cell cycle, RNA processing and protein localization related bioprocesses, were mainly covered by EHF enriched GO BPs, which also reflected the functional relations between EHFs and VTPs.

Altogether, these results highly indicate the importance of both EHFs and VTPs in real infections, and their potential regulating or coordinating associations.

EHFs and VTPs in human PPI network

It is intriguing that EHFs and VTPs show consistency in regulating orientation and to some extent are functionally related. Therefrom, we sought to explore the functional associations of EHFs and VTPs using the experimentally validated protein-protein interaction data.

Protein-protein interaction (PPI) represents a pivotal aspect of protein function in numerous physiological and pathological processes. Networks of protein interactions can reveal how complex molecular processes are activated in the cell, and visualization of the human interactome has been made possible where each point represents a protein and each line between them is an interaction[9]. To give an in-depth and global view of relationship between EHFs and VTPs at translational level, we first investigate the topological properties of EHF protein products and VTPs in the human PPI network. The human PPI network consists of 39,204 interactions among 9,673 human proteins[13]. In general, a total of 487 EHFs (59.3%) and 690 VTPs (82.4%) for HIV-1, as well as 894 EHFs (73.4%) and 1214 VTPs (99.7%) for IAV were mapped onto the human PPI network (Fig. 4A and B).

We applied multiple metrics, including node degree, betweenness centrality, closeness centrality, and *k*-core, to evaluate topological properties of EHF and VTP in the human PPI network for HIV-1 and IAV, respectively. Meanwhile, we created a control network based on 10^6 permutations. As a result, we found that for both HIV-1 and IAV, the average node degree of EHFs (i.e., 20.4 for HIV-1 and 17.6 for IAV) and VTPs (i.e., 21.2 for HIV-1 and 17.4 for IAV) are significantly higher than control (P<0.001, Table S4). This suggested that EHFs and VTPs tend to occupy the centre of the network. However, the calculated betweenness centrality and closeness centrality for EHFs and VTPs of both viruses are of no significance (P>0.001, Table S4). This indicates that EHFs and VTPs are not the crucial bridging factors in the human PPI network.

K-core is one of the anatomical measurements for structural backbone of a network, defined as the maximal connected subnetwork of it in which degree of all nodes is not less than *k*. It has been widely used to describe the clustering structure of a network or the evolution of a random network[14–18]. We calculated the percentages of EHFs and VTPs as positioned at each structural level of human PPI network for the two viruses respectively. Obviously, more than 50% EHFs and VTPs of both viruses reside within the 0 to 2 level of coreness in the human PPI network (Fig. 4C). The average coreness of EHFs and VTPs for HIV-1 are 2.27 and 3.20, and those for IAV are 2.77 and 2.50. We conducted a randomization test and found that all average coreness of EHFs and VTPs for the two viruses were significantly higher than control (P < 0.005 for all). The average coreness of VTPs was higher than that of EHFs for HIV-1. However, the result was without significant differences (P = 0.06). For IAV, those values are almost the same (P = 0.24). Altogether, these results again indicate that both EHFs and VTPs tend to locate at the global centre of human PPI network that holds regulating potential to extensive proteins.

To further reveal the interacting relationships between EHFs and VTPs, we explored the distributional enrichment among the highly interconnected modules identified in the human PPI network. As a result, we found 18 modules containing 10 or more nodes, in which VTPs of HIV-1 were enriched in 7 modules, while EHFs of HIV-1 were not enriched in any module (Table S5). For IAV, EHFs and VTPs were enriched in 6 and 1 module(s), respectively (Table S5). Of note, there is no module exclusively enriched of EHFs or VTPs for both HIV-1 and IAV. This again highlights the extensive functional associations between EHFs and VTPs.

Next, we tried to further investigate the distribution relationship between EHFs and VTPs, respectively. To this end, we calculated the shortest distance between any EHF-EHF and VTP-VTP pair in human PPI network for HIV-1 and IAV, and presented as distances with respect to percentages. The average internal mean distance of EHFs and VTPs for HIV-1 are 4.00 (P = 0.0001) and 3.63 (P<10^{-5}), and that for IAV are 3.76 and 3.78 (P <10^{-5} for all) (Fig. 4D). Notably, for both viruses, more than 70% of VTPs are within the distances ranging from 0 to 4, whereas more than 50% of EHFs are within the distances ranging from 4 to 8. It can be inferred that compared to the protein products of EHFs, VTPs are more of direct interacting or regulating relationships.

To explore the associations of VTPs and EHFs in human PPI network, we further calculated the distances of EHFs from their nearest VTPs and that of VTPs from their nearest EHFs for the two viruses (Fig. 5). We found that, with EHFs being the benchmark, the largest proportion of EHFs for HIV-1 and IAV (46. 01% and 49.66%) are of 1-step distance to their nearest VTPs (Fig. 5A), which are of statistical significance (P < 0.001). These results demonstrate that a large proportion of EHFs may

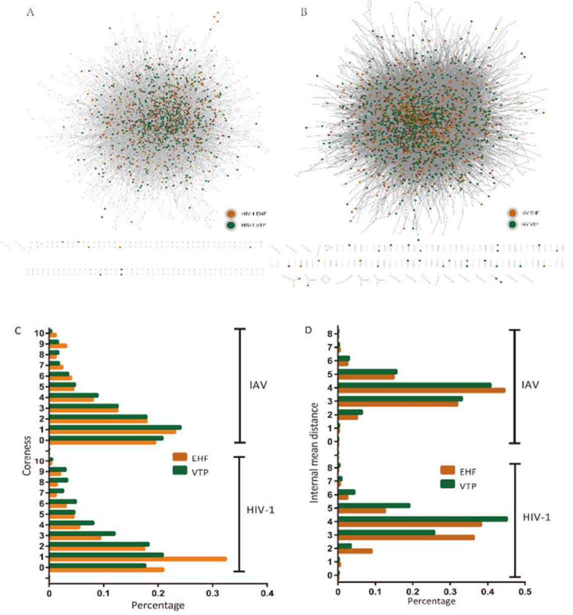


Fig. 4 Topological properties of essential host factors (EHFs) and virus targeting proteins (VTPs) for human immunodeficiency virus type 1 (HIV-1) and influenza A virus (IAV) in human protein-protein interaction (PPI) network. Cytoscape overview on distributions of EHFs and VTPs of (a) HIV-1 and (b) IAV in human PPI network. Statistics of (c) coreness and (d) internal mean distance distribution as calculated in percentage.

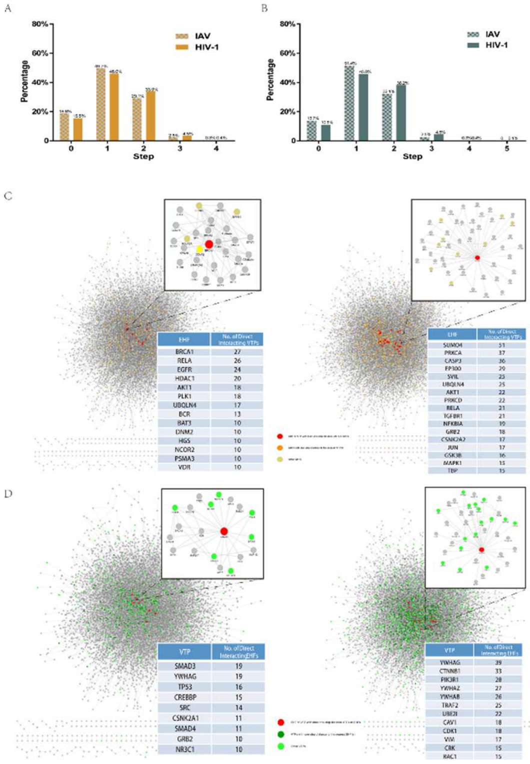


Fig. 5 Distributional association of virus targeting proteins (VTPs) and essential host factors (EHFs) for human immunodeficiency virus type 1 (HIV-1) and influenza A virus (IAV) in human protein-protein interaction (PPI) network. Statistics on distance distribution of the number of nearest (a) VTPs and (b) EHFs as calculated with EHFs or VTPs as benchmarks. The number of VTPs/EHFs with different calculated steps from its nearest EHF/VTP is presented as percentage in total VTPs/EHFs. Cytoscape overview of (c) VTPs with one-step distance to the nearest EHFs and (d) EHFs with one-step distance to the nearest VTPs for HIV-1 and IAV, as in the subnetwork of human PPI network that shows exclusively PPI relations of EHFs and VTPs. For each case, VTPs/EHFs with 10 or more one-step distance EHFs/VTPs are specifically listed (Please note that for IAV, only EHFs/VTPs with 15 or more one-step distance VTPs/EHFs are listed). And the VTP/EHF with most one-step distance EHFs/VTPs is highlighted in the subnetwork and its connections in an enlarged version are shown independently.

perform direct regulation to VTPs, which better explains their substantial indispensability for viral infection in loss of function experiment like RNAi.

Meanwhile, with VTPs being the benchmark, we found that the longest distance between a pair of VTP and EHF for HIV-1 and IAV is 5 and 4 steps, respectively (Fig. 5B). Of note, for HIV-1 and IAV respectively, 10.83% and 13.70% of their VTPs coincide with their nearest EHFs in human PPI network (Fig. 5B). And 45.83% and 51.38% of their VTPs are at only one step away from their nearest EHFs for both HIV-1 and IAV (Fig. 5B), which are of statistical significance ($P < 0.001$). These results highly indicated the economical and efficient viral exploitation of host proteins during infection, because these VTPs that viruses directly interact with may influence multiple EHFs to co-ordinately fulfil the intracellular survival and persistent infection.

Further, we identified the top ranking EHFs (or VTPs) that show extensive direct interactions (i.e., one-step distance) with VTPs (or EHFs) in the human PPI network for HIV-1 and IAV, respectively (Fig. 5C and D; detailed information is provided in Table S6). For HIV-1, 123 out of the 487 mapped EHFs show direct interactions with 2 or more VTPs, and 14 EHFs show direct interactions with 10 or more VTPs, including BRCA1, RELA, EGFR, HDAC1, AKT1, PLK1, UBQLN4, BCR, BAT3, DNM2, HGS, NCOR2, PSMA3, VDR (Fig. 5C, left). And BRCA1(breast cancer 1, early onset), which encodes a nuclear phosphoprotein that plays a role in maintaining genomic stability, and also acts as a tumor suppressor, is the EHF whose gene product shares most extensive direct interactions with VTPs (i.e., 27). Notably, it has been reported that BRCA1 plays an important role for viral transcription in HIV-1 infection and phosphorylate of BRCA1 could be used as a potential host-based therapy for combined treatment of HIV-1 infection[19].

Likewise, 182 out of the 690 mapped VTPs show direct interactions with 2 or more EHFs, and notably 9 VTPs show direct interactions with 10 or more EHFs, including SMAD3, YWHAG, TP53, CREBBP, SRC, CSNK2A1, SMAD4, GRB2, NR3C1(Fig. 5D, left). SMAD3, which functions as a transcriptional modulator activated by transforming growth factor-beta (TGF- beta) and is thought to play a role in the regulation of carcinogenesis, is the VTP that shares most extensive direct interactions (i.e., 19) with EHF gene products. Yamaguchi etc. has proved that shift of SMAD3 phospho-isoform signaling from tumor suppression to carcinogenesis increases the risk of hepatocellular carcinoma in hepatitis C virus (HCV) infection[20].

For IAV, 271 out of the 894 mapped EHFs show direct interactions with 2 or more VTPs, and 10 EHFs show direct interactions with 20 or more VTPs, including SUMO4, PRKCA, CASP3, EP300, SVIL, UBQLN4, AKT1, PRKCD, RELA, TGFBR1 (Fig. 5C, right). SUMO4 (Small ubiquitin-related modifier 4) is the EHF whose gene product shares most extensive direct interactions (i.e., 51) with VTPs. It locates in the cytoplasm and specifically modifies IKBA, leading to negative regulation of NF-kappa-B-dependent transcription of the IL12B gene. A specific polymorphism in SUMO4 leading to the M55V substitution, has been associated with type I diabetes[21]. In addition, The ataxin-1 ubiquitin-like interacting protein (UBQLN4) has been identified as a host cellular target for the small hydrophobic protein of mumps virus[22].

Likewise, 361 out of the 1214 mapped VTPs showed direct interactions with 2 or more EHFs, and 12 VTPs showed direct interactions with 15 or more EHFs, including YWHAG, CTNNB1, PIK3R1, YWHAZ, YWHAB, TRAF2, UBE2I, CAV1, CDK1, VIM, CRK, RAC1 (Fig. 5D right). YWHAG (Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, gamma), is a member of 14-3-3 protein family which mediate signal transduction by binding to phosphoserine-containing proteins. YWHAZ is the VTP that shares most extensive direct interactions (i.e., 39) with EHF gene products, and it has been suggested that YWHAZ plays an important role in muscle tissue. And it has also been reported that during the course of IAV infection, GSK-3β-mediated β-catenin degradation in adherent junctions leads to vascular hyperpermeability[23].

Together, these results suggested that EHF and VTP are intensively associated functional sets in viral infection, and those most extensively interacting components may provide new insights into the differential features of specific viral infection and also serve as promising candidate targets for developing host-directed antiviral drugs.

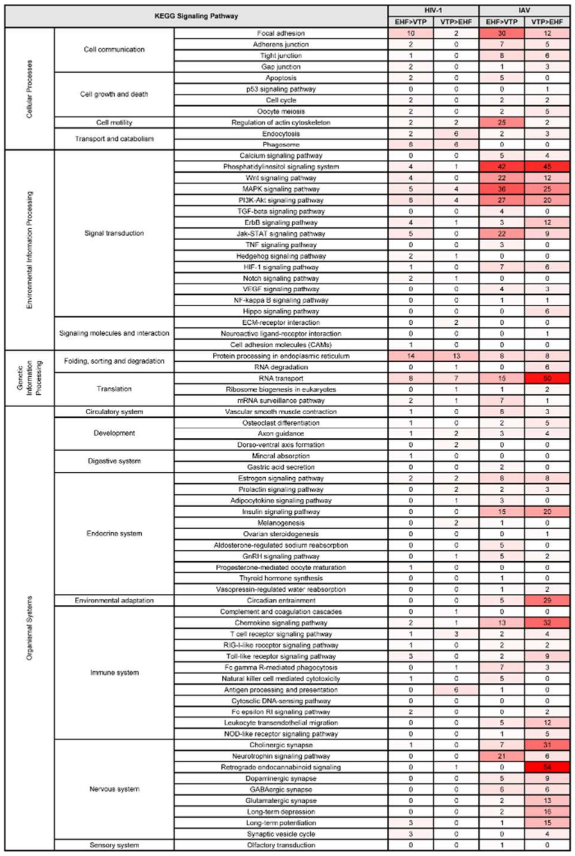
EHFs and VTPs in KEGG pathways

Protein interaction networks encode a variety of signaling processes that occur in the cell, in which signaling pathways are subnetworks of proteins that communicate via a series of interactions and are often only activated under specific conditions. To further explore the potential associations between EHFs and VTPs as functional coordinators, we mapped EHFs and VTPs of HIV-1 and IAV onto KEGG pathways. In general, 33.01% (271/821) EHFs and 50.66% (424/837) VTPs of HIV-1, as well as 44.83% (546/1218) of EHFs and 39.66% (629/1586) VTPs of IAV have been mapped onto KEGG signaling pathways, respectively (detailed information is provided in Table S7).

Signaling pathways typically contain upstream proteins that sense changes in the environment or that are directly involved in host-pathogen interactions. These proteins thereafter trigger a signaling cascade leading to downstream transcription factors, which consequently elicit regulatory programs. Thus, we respectively calculated the EHF-VTP pairs in the same pathways for HIV-1 and IAV, and further marked their positions to examine possible relation patterns.

Through functional annotation and overlap analysis, we identified 191 pairs of EHF and VTP for HIV-1 positioning in the same pathways (Fig. 6, detailed information is provided in Table S8). Notably, 113 (59.16%) pairs showed the relation pattern that EHF is the upstream regulator of VTP (Hypergeometric test, $P < 10^{-5}$). Likewise, 981 pairs of EHF and VTP for IAV resided in the same pathways, and intriguingly, 431 (43.93%) pairs also presented the relation pattern that

EHF is the upstream regulator of VTP (Hypergeometric test, $P < 10^{-5}$).



KEGG Signaling Pathway			HIV-1		IAV	
			EHF>VTP	VTP>EHF	EHF>VTP	VTP>EHF
Cellular Processes	Cell communication	Focal adhesion	10	2	30	12
		Adherens junction	2	0	7	5
		Tight junction	1	0	8	6
		Gap junction	2	0	1	3
	Cell growth and death	Apoptosis	2	0	5	0
		p53 signaling pathway	0	0	0	1
		Cell cycle	2	0	2	2
		Oocyte meiosis	2	0	2	5
	Cell motility	Regulation of actin cytoskeleton	2	2	25	2
	Transport and catabolism	Endocytosis	2	6	2	3
		Phagosome	6	6	0	0
Environmental Information Processing	Signal transduction	Calcium signaling pathway	0	0	5	4
		Phosphatidylinositol signaling system	4	1	42	45
		Wnt signaling pathway	4	0	22	12
		MAPK signaling pathway	5	4	36	25
		PI3K-Akt signaling pathway	6	4	27	20
		TGF-beta signaling pathway	0	0	4	0
		ErbB signaling pathway	4	1	3	12
		Jak-STAT signaling pathway	5	0	22	9
		TNF signaling pathway	0	0	3	0
		Hedgehog signaling pathway	2	1	0	0
		HIF-1 signaling pathway	1	0	7	6
		Notch signaling pathway	2	1	0	0
		VEGF signaling pathway	0	0	4	3
		NF-kappa B signaling pathway	0	0	1	1
		Hippo signaling pathway	0	0	0	6
	Signaling molecules and interaction	ECM-receptor interaction	0	2	0	0
		Neuroactive ligand-receptor interaction	0	0	0	1
		Cell adhesion molecules (CAMs)	1	0	0	0
Genetic Information Processing	Folding, sorting and degradation	Protein processing in endoplasmic reticulum	14	13	8	8
		RNA degradation	0	1	0	6
	Translation	RNA transport	8	7	15	50
		Ribosome biogenesis in eukaryotes	0	1	1	2
		mRNA surveillance pathway	2	1	7	1
Organismal Systems	Circulatory system	Vascular smooth muscle contraction	1	0	6	3
	Development	Osteoclast differentiation	1	0	2	5
		Axon guidance	1	2	3	4
		Dorso-ventral axis formation	0	2	0	0
	Digestive system	Mineral absorption	1	0	0	0
		Gastric acid secretion	0	0	2	0
	Endocrine system	Estrogen signaling pathway	2	2	6	8
		Prolactin signaling pathway	0	2	2	3
		Adipocytokine signaling pathway	0	1	3	0
		Insulin signaling pathway	0	0	15	20
		Melanogenesis	0	2	1	0
		Ovarian steroidogenesis	0	0	0	1
		Aldosterone-regulated sodium reabsorption	0	0	5	0
		GnRH signaling pathway	0	1	5	2
		Progesterone-mediated oocyte maturation	1	0	0	0
		Thyroid hormone synthesis	0	0	1	0
		Vasopressin-regulated water reabsorption	0	0	1	2
	Environmental adaptation	Circadian entrainment	0	0	5	29
	Immune system	Complement and coagulation cascades	0	1	0	0
		Chemokine signaling pathway	2	1	13	32
		T cell receptor signaling pathway	1	3	2	4
		RIG-I-like receptor signaling pathway	1	0	2	2
		Toll-like receptor signaling pathway	3	0	2	9
		Fc gamma R-mediated phagocytosis	0	1	7	3
		Natural killer cell mediated cytotoxicity	1	0	5	0
		Antigen processing and presentation	0	6	1	0
		Cytosolic DNA-sensing pathway	0	0	0	0
		Fc epsilon RI signaling pathway	2	0	0	2
		Leukocyte transendothelial migration	0	0	5	12
		NOD-like receptor signaling pathway	0	0	1	5
	Nervous system	Cholinergic synapse	1	0	7	31
		Neurotrophin signaling pathway	0	0	21	6
		Retrograde endocannabinoid signaling	0	1	0	54
		Dopaminergic synapse	0	0	5	9
		GABAergic synapse	0	0	6	6
		Glutamatergic synapse	0	0	2	13
		Long-term depression	0	0	2	16
		Long-term potentiation	3	0	1	15
		Synaptic vesicle cycle	3	0	0	4
	Sensory system	Olfactory transduction	0	0	1	0

Fig. 6 Categorized statistical list of essential host factors (EHFs) and virus targeting proteins (VTPs) pairs as mapped in the same Kyoto Encyclopedia of Genes and Genomes (KEGG) signaling pathways for human immunodeficiency virus type 1 (HIV-1) and influenza A virus (IAV). Squares are differentially colored according to the value ranges. Note: EHF>VTP, EHF locates at the upstream of VTP in the same KEGG signaling pathways; VTP>EHF, VTP locates at the upstream of EHF in the same KEGG signaling pathways.

In total, we found the mapped EHF-VTP pairs showing membership in 50 and 64 different KEGG signaling pathways for HIV-1 and IAV, respectively (Fig. 6). 62 of the total 74 non-repeated KEGG signaling pathways exclusively presented the relation pattern that EHF is the upstream regulator of VTP. Specifically, 9 of these 62 pathways were only observed for HIV-1, 33 were only observed for IAV, and 20 were observed for both HIV-1 and IAV.

Notably, for signaling pathway(s) functionally categorized as in cell communication (e.g., focal adhesion, Fig. 7, detailed information is provided in Table S9), cell growth and death (e.g., apoptosis, p53 signaling pathway, and cell cycle), cell motility (i.e., regulation of actin cytoskeleton), signal transduction (e.g., Wnt signaling pathway), and immune system (e.g., chemokine signaling pathway), the relation pattern that EHF is the upstream regulator of VTP takes predominantly advantageous proportions for both HIV-1 and IAV. In particular, focal adhesion is the top ranking pathway that holds the most EHF-upstream pairs and the least VTP-upstream pairs for both viruses. Focal adhesions play significant roles in important biological processes such as cell proliferation, cell differentiation, regulation of gene expression and cell survival[24,25]. Interestingly, when viewed in an integrative manner, it shows cross-talk with other EHF-upstream predominant pathways with high rankings (Fig. 7), e.g., Wnt signaling pathway, MAPK signaling pathway and PI3K-Akt signaling pathway.

Collectively, these results reveal the potential regulating relations between EHFs and VTPs, and further emphasize the essential role of EHFs in gene expression regulation.

Discussion

As one of the most promising high-throughput approaches, genome-wide RNAi screen has provided abundant host genes as essential host factors for different viral infections. Although the biological mechanism underlying the experiment itself is a reflection of the essentiality of these genes, the low overlap rates among EHF results from multiple screens on the same virus have greatly compromised the fundamental generalization of this particular approach and the results (Table S10). Here, through meta-analysis on the associations between EHFs and VTPs of HIV-1 and IAV as indicative models, we made the first attempt to reveal the common and differential features of EHFs and VTPs as important to viral infection at both transcriptional and translational levels, and notably, the significant regulating roles of EHFs.

To minimize the inevitable noise in genome-wide RNAi screen results, we used genes classified as confirmed hits instead of hits to ensure that these gene candidates are with lower false positive rates. For better recovery of generic features of EHFs for viral infections, we computationally combined confirmed hits from multiple genome-wide RNAi screens regarding a specific viral species, and designated this gene set as EHFs for the virus. The major findings, including limited overlap, consistent regulation orientation in HTR to infection, similar topological properties in human PPI network,

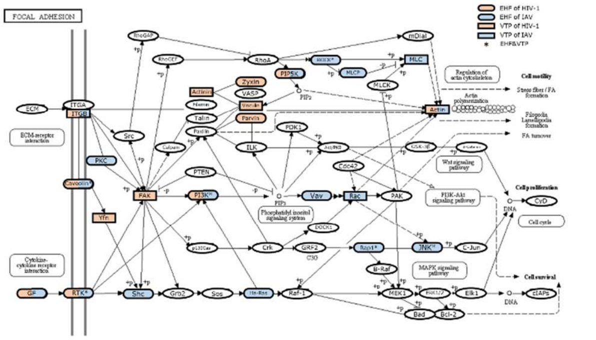


Fig. 7 Detailed information of distribution pattern of essential host factors (EHFs) and virus targeting proteins (VTPs) for human immunodeficiency virus type 1 (HIV-1) and influenza A virus (IAV) in Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway of focal adhesions.

and regulation relations in KEGG pathways, showed high consistency for both HIV-1 and IAV. Therefore, these findings concerning associations of EHFs and VTPs at multiple perspectives might provide a general conclusion that integrates the diverse results of host genes for viral infection from different approaches. Meanwhile, there are also virus-specific features identified in individual analysis item, e.g., regulation orientation as compared to HTR, and the same KEGG signaling pathway that holds converse EHF-VTP regulating patterns. These are worthy of further investigation if individual viral infection is the study focus. Overall, our findings further strengthen the efficiency of viral exploitation of host factors, and offer new evidence to the biological importance of EHFs in a systematic way. Moreover, based on the extensive direct interactions between the functional sets of EHF and VTP, it may also shed light on the identification of targets for host-directed antiviral drugs.

Hopefully, meta-analysis like this can be performed at larger scales, if EHF and VTP data for many other viral types are available and/or validated from different resources. It can be expected that, combined with experimental evidence, more feature rich association may provide valuable insight into the complex virus-host interactions and the exquisite regulating patterns among different types of host factors, and finally benefit the development of antiviral therapeutics.

Conclusions

In this study, we investigated the functional associations of EHFs and VTPs for HIV-1 and IAV in a systematic manner. At transcriptional level, both EHFs and VTPs of the two viruses are among the most differentially dysregulated genes in host transcriptional response of viral infection although the overlap between them is limited. At translational level, EHFs and VTPs (1) tend to locate at topological important positions, i.e., global centre of the human PPI network and show extensive direct interactions; (2) most EHFs are at the upstream of VTPs in the same signaling pathways as regulators of specific biological processes. Our analysis presented multi-level close functional associations of the two important host factor types and emphasized the regulating significance of EHFs for viral infection.

Materials and Methods

Essential host factors (EHFs) for viral infection

Through literature search, we collected before April 1st, 2014, four and six publications (Table S1) reporting genome-wide RNAi screens that experimentally identified EHFs for infections of HIV-1 and IAV, respectively. After comparison, we found that they are the two viruses that received the most extensive genome-wide RNAi screens. Despite multi-level validation, the overlap of the confirmed hits between different screens conducted by various laboratories for the same virus is limited (detailed information is provided in Table S1). Therefore, we computationally combined the results of confirmed hits after multi-level screens in each study (the workflow is illustrated in Fig. 1). Of note, we took account of only EHFs with phenotype of decreased infection when silenced *in vitro*.

VTPs and human protein-protein interaction (PPI) information

The virus targeting proteins (VTPs) of HIV-1 and IAV were collected from the VirHostNet database (renewed to December 18th, 2014), which provides complete, manually curated and up-to-date information on protein-protein interactions between viruses and hosts integrated from 9 related public databases[4]. The VirHostNet database provided information on virus-host interactions confirmed by 77 test methods including two hybrid, co-immunoprecipitation and tandem affinity purification. A total of 2054 VTPs of HIV-1 and IAV mainly detected by co-immunoprecipitation were reported in the database. The human PPI information used in this study was downloaded from the Human Protein Reference Database (HPRD) (Release 9)[13], which to date includes 9,673 human proteins and 39,204 interactions.

Host transcriptional response to viral infection

The genome-scale gene expression profiles representing HTR to HIV-1 and IAV infections were collected from the Gene Expression Omnibus (GEO) database[26]. The profile data we used in this study should be generated using HG-U133A cartridge arrays or HG-U133 Plus 2.0, and comply with the following principles: (I) The project should contain at least one sample of untreated specific pathogen infection with infectious disease state or *in vitro* infection for at least one hour; (II) The project should contain at least one sample of control (It could be uninfected, mock-infected, healthy control or other blank controls defined by submitters); (III) The project would be discarded if more than ten-percent of probes miss data values in its series matrix file. And the strain type and human cell type should match those used in the corresponding screens for EHFs.

The GSEA desktop application[27] was used to estimate the regulation tendency of EHFs and VTPs as in host gene expression profiles upon infection. We used its default parameters to generate the gene rank list for each profile. And classic scoring scheme was applied to get enrichment scores for EHFs and VTPs. Moreover, one thousand permutations were generated to estimate the significance of their enrichment scores.

Topological analysis of EHFs and VTPs in human PPI network

We respectively calculated coreness and internal mean distance for EHFs and VTPs of HIV-1 and IAV in the human PPI network. The *K*-core of a network is defined as the maximum sub-network obtained by pruning all nodes with a degree lower than *k*, and higher coreness nodes belong to higher levels of the core sub-network. Internal mean distance in a network measures dispersion or degree of scatter of a group of nodes in human PPI network.

The human PPI network based on Human Protein Reference Database (HPRD) was used as background control to determine whether the corenesses of EHFs and VTPs of HIV-1 or IAV are significantly higher. All nodes in the human PPI network were shuffled 10^6 times while calculating and preserving their average *K*-cores. Let X represent the actual average *K*-core of EHFs or VTPs of a designated virus, the *P*-value was computed as the ratio of the actual number larger than X to 10^6.

To examine whether EHFs of HIV-1 or IAV preferentially distribute near their VTPs (or VTPs of these two viruses preferentially distribute near their EHFs) in the human PPI network, 39,204 pairs of interactions obtained from the HPRD database were used as background control. First, we computed the actual average distance of EHFs from their nearest VTPs (or VTPs from their nearest EHFs). Next, by keeping unchanged number of both EHFs and VTPs, we randomly chose a group of protein nodes of the same number in human PPI network for control. In total, we generated 10^6 independent randomized samples to calculate the statistical significations.

PPI module identification and enrichment analysis

We use Cytoscape plugin MCODE[16] to find highly connected modules in the human PPI network from HPRD[13]. Gene enrichment analysis is performed for the VTPs and EHFs based on hypergeometric distribution with phyper function in R.

Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways analysis

We also used KEGG database (including 118 human pathways relating to 3306 genes) to explore the associations between EHFs and VTPs as genes in signaling pathways. To test if EHFs tend to locate at the upstream of VTPs in signaling pathways, we calculated the actual proportion of three circumstances, i.e., EHF locates at the upstream of VTP, EHF locates at the downstream of VTP, EHF and VTP are not in a same pathway. We then sampled the same number of node pairs from KEGG for 10^6 times and calculated the *P*-value.

Competing financial interests

The authors declare no competing financial interests.

Author contributions

X.B., H.B. and S.W. formulated the idea of the paper and supervised the research. D.X., L.H., Y.L., Y.L., S.H. and H.B. performed the research and analysed the data. D.X., L.H., H.B. and X.B. wrote the paper.

Acknowledgements

Financial support from the National Nature Science Foundation of China (U1435222 81273488, 81230089), National Key Technologies R&D Program for New Drugs (2012ZX09301-003), Program of International S&T Cooperation (2014DFB30020) and National Science & Technology Pillar Program of China (2012BAI29B07).

Notes and references

1. Karlas, A. *et al.* Genome-wide RNAi screen identifies human host factors crucial for influenza virus replication. *Nature* **463,** 818–822 (2010).
2. Law, G. L., Korth, M. J., Benecke, A. G. & Katze, M. G. Systems virology: host-directed approaches to viral pathogenesis and drug targeting. *Nat. Rev. Microbiol.* **11,** 455–466 (2013).
3. Panda, D. & Cherry, S. Cell-based genomic screening: elucidating virus-host interactions. *Curr Opin Virol* **2,** 784–792 (2012).
4. Navratil, V. *et al.* VirHostNet: a knowledge base for the management and the analysis of proteome-wide virus-host interaction networks. *Nucleic Acids Res.* **37,** D661–668 (2009).
5. Kumar, R. & Nanduri, B. HPIDB--a unified resource for host-pathogen interactions. *BMC Bioinformatics* **11 Suppl 6,** S16 (2010).
6. Fu, W. *et al.* Human immunodeficiency virus type 1, human protein interaction database at NCBI. *Nucleic Acids Res.* **37,** D417–422 (2009).
7. Liu, Y. *et al.* EHFPI: a database and analysis resource of essential host factors for pathogenic infection. *Nucleic Acids Res.* **43,** D946–955 (2015).
8. Sedaghat, A. R. *et al.* Chronic CD4+ T-cell activation and depletion in human immunodeficiency virus type 1 infection: type I interferon-mediated disruption of T-cell dynamics. *J. Virol.* **82,** 1870–1883 (2008).
9. Sutejo, R. *et al.* Activation of type I and III interferon signalling pathways occurs in lung epithelial cells infected with low pathogenic avian influenza viruses. *PLoS ONE* **7,** e33732 (2012).
10. Okumura, A., Lu, G., Pitha-Rowe, I. & Pitha, P. M. Innate antiviral response targets HIV-1 release by the induction of ubiquitin-like protein ISG15. *Proc. Natl. Acad. Sci. U.S.A.* **103,** 1440–1445 (2006).
11. Yu, X. *et al.* Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex. *Science* **302,** 1056–1060 (2003).
12. Planz, O. Influenza viruses and intracellular signalling pathways. *Berl. Munch. Tierarztl. Wochenschr.* **119,** 101–111 (2006).
13. Keshava Prasad, T. S. *et al.* Human Protein Reference Database--2009 update. *Nucleic Acids Res.* **37,** D767–772 (2009).
14. Seidman, S. B. Network structure and minimum degree. *Social Networks* **5,** 269–287 (1983).
15. Dorogovtsev, S. N., Goltsev, A. V. & Mendes, J. F. F. k-core organization of complex networks. *Physical Review Letters* **96,** (2006).
16. Bader, G. D. & Hogue, C. W. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* **4,** 2 (2003).
17. Wuchty, S. & Almaas, E. Peeling the yeast protein network. *Proteomics* **5,** 444–449 (2005).

18. Alvarez-Hamelin, J. I., Dall'Asta, L., Barrat, A. & Vespignani, A. k-core decomposition: a tool for the visualization of large scale networks. *arXiv:cs/0504107* (2005).
19. Guendel, I. *et al.* BRCA1 functions as a novel transcriptional cofactor in HIV-1 infection. *Virol. J.* **12,** 40 (2015).
20. Yamaguchi, T., Yoshida, K., Murata, M. & Matsuzaki, K. Smad3 phospho-isoform signaling in hepatitis C virus-related chronic liver diseases. *World J. Gastroenterol.* **20,** 12381–12390 (2014).
21. Song, G. G., Choi, S. J., Ji, J. D. & Lee, Y. H. Association between the SUMO4 M55V (A163G) polymorphism and susceptibility to type 1 diabetes: a meta-analysis. *Hum. Immunol.* **73,** 1055–1059 (2012).
22. Woznik, M. *et al.* Mumps virus small hydrophobic protein targets ataxin-1 ubiquitin-like interacting protein (ubiquilin 4). *J. Gen. Virol.* **91,** 2773–2781 (2010).
23. Hiyoshi, M. *et al.* Influenza A virus infection of vascular endothelial cells induces GSK-3β-mediated β-catenin degradation in adherens junctions, with a resultant increase in membrane permeability. *Arch. Virol.* **160,** 225–234 (2015).
24. Kanehisa, M. *et al.* Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Research* **42,** D199–D205 (2014).
25. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28,** 27–30 (2000).
26. Barrett, T. *et al.* NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res.* **41,** D991–995 (2013).
27. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **102,** 15545–15550 (2005).