Review

Cell PRESS

*Microbial systems biology*

# Virus–host interactomes and global models of virus-infected cells

**Caroline C. Friedel[1,2] and Jürgen Haas[3,4]**

[1] Institut für Pharmazie und Molekulare Biotechnologie, Universität Heidelberg, 69120 Heidelberg, Germany
[2] Institut für Informatik, Ludwig-Maximilians-Universität München, 80333 München, Germany
[3] Division of Pathway Medicine and Edinburgh Infectious Diseases, University of Edinburgh, Edinburgh EH16 4SB, UK
[4] Max von Pettenkofer Institut, Ludwig-Maximilians-Universität München, 80336 München, Germany

**Novel high-throughput technologies such as yeast two-hybrid and RNA interference (RNAi) screens provide the tools to study interactions between viral proteins and the host on a genomic scale. In this review, we provide an overview of studies in which these technologies were applied and of computational approaches for the analysis of the identified viral interactors in the context of the host cell. The results of these studies illustrate the advantages of integrative systems biology approaches in the investigation of viral pathogens.**

## Systems biology-based approaches to identify potential antiviral drug targets

Viral pathogens such as HIV, hepatitis and influenza virus constitute a major threat to human health. At the end of 2009, 33.3 million people were living with HIV and an estimated 1.8 million AIDS-related deaths were reported in that year [1]. Infection rates are even higher for hepatitis viruses: 130–170 million people are chronically infected with hepatitis C virus (HCV) and 350 million with hepatitis B (HBV) (WHO fact sheets, http://www.who.int/mediacentre/factsheets/). Influenza, which even in a normal season causes 250 000–300 000 deaths worldwide, has been responsible for four pandemics since the beginning of the 20th century. The most lethal of these was the 1918 'Spanish' influenza pandemic, which killed an estimated 50–100 million people [2]. The rapid spread of viruses, in particular by air travel, during the more recent SARS outbreak and 'swine flu' pandemic demonstrated the vulnerability of modern society to infectious diseases. Furthermore, viruses are also associated with a significant number of malignancies in humans. Epstein-Barr virus (EBV), which has an estimated infection rate of 90% of the adult human population [3], is linked to several tumors, such as Burkitt's lymphoma, nasopharyngeal carcinoma and Hodgkin's disease [4]. Together with HBV, HCV, human papilloma virus, human T-cell lymphotropic virus and Kaposi's sarcoma-associated herpesvirus (KSHV), EBV is involved in the formation of an estimated 10–15% of cancers [5].

Effective antiviral drugs are currently only available against a few viruses. Because licensed antiviral drugs all target viral proteins, they are usually virus-specific and vulnerable to the fast development of resistance, particularly in the case of RNA viruses, which possess high mutation rates. Drugs that target virus–host protein interactions, however, might have a broader spectrum and be more stable with regard to the development of resistance. To this end, novel genome-scale screening technologies allow the comprehensive analysis of host factors involved in virus infection, and thus help select potential targets for antiviral drug therapy. In this review, we will discuss the outcome of such screens as well as the bioinformatic tools developed to analyze the data generated by these approaches.

## Identification of host factors involved in virus infection

Originally, protein interactions between viral proteins and the corresponding host interaction partners were only analyzed on a small scale. Even so, when Dyer *et al.* [6] compiled a list of pathogen–host interactions from public databases in 2007, they were able to collect approximately 10 000 virus–host interactions. The vast majority of these (~80%), however, comprised direct and indirect interactions of HIV with its human host and for most other viruses only a small number of virus–host interactions were described. Because HIV-1 only encodes 15 distinct proteins, the number of interactions for each protein appears to be exorbitantly high. Although a manual survey of the available literature on HIV–host protein interactions in 2008 identified only 2589 interactions [7], this still implies an average of approximately 170 physical or functional interactions per protein. Thus, although more stringent criteria were applied in the manual survey than were required for the database deposits, it is still questionable whether all of these 2589 HIV-1–human interactions are truly specific. Subsequent to these surveys, a number of databases have been published specifically for virus–host interactions, including the HIV-1, human protein interaction database at NCBI [8], VirHostNet [9], VirusMINT [10], PIG [11] and HPIDB [12]. Whereas the first three databases are based on extensive literature curation, the last two mostly draw from other public resources and probably contain much more spurious and unverified interactions. To date, none of these databases has established itself as a standard repository for virus–host interactions with the exception of the NCBI HIV-1 database for this specific virus.

*Corresponding author:* Haas, J. (juergen.haas@ed.ac.uk)

0966-842X/$ – see front matter  doi:10.1016/j.tim.2011.07.003 Trends in Microbiology, October 2011, Vol. 19, No. 10 

**Table 1. Overview of large-scale Y2H screens of virus–host interactions**

| Virus species | cDNA library | Number of interacting proteins | Number of interactions | Refs |
|---|---|---|---|---|
| Epstein-Barr virus (EBV) | Human spleen cDNA library ($10^5$–$10^6$ proteins) | 112 | 173 | [19] |
| Hepatitis C virus (HCV) NS5A protein | HeLa cDNA library | 3 | 3 | [70] |
| Hepatitis C virus (HCV) | Fetal brain and spleen cDNA libraries | 278 | 314 | [20] |
| HIV-1 | Human leukocyte cDNA library | 8 | 9 | [23] |
| Moloney murine leukemia virus (MoMLV) integrase | Mouse WEHI-3B cDNA library, mouse T-cell cDNA library | 121 | 121 | [24] |
| HIV-1 integrase | 27 MoMLV interacting proteins | 12 | 12 | [24] |
| HIV-1 integrase | Human CEMC7 cell CDNA library | 4 | 4 | [25] |
| HIV-1 integrase | Human lymphocytes cDNA library | 13 | 13 | [26] |
| Vaccinia virus | Four libraries: human tongue/tonsil, human spleen, breast tumor and prostate tumor cell lines, NIH Mammalian Gene Collection (MGC) | Not specified | 109 | [21] |
| Influenza (H1N1) | Human ORFeome v3.1 collection (~12 000 proteins) | 87 | 135 | [22] |
| Influenza (H3N2) | Human ORFeome v3.1 collection (~12 000 proteins) | 66 | 81 | [22] |
| Equine infectious anemia virus (EIAV) S2 protein | Human spleen cDNA library | 2 | 2 | [29] |

As a result of the small genome size of viruses, genome-wide screens of direct intraviral protein interactions with the yeast two-hybrid system (Y2H) have been performed for several viruses [13–18]. Because virus–host screens require a much larger number of interactions to be tested, the first genome-wide Y2H screen of virus–host interactions was only published for EBV in 2007 by Calderwood *et al.*, who identified 173 EBV–human interactions [19]. Since then, genome-scale virus–host Y2H studies identified 314 virus–human interactions for HCV [20], 109 interactions for vaccinia virus [21], 135 and 81 interactions for H1N1 and H3N2 influenza, respectively [22], and nine interactions for HIV-1 [23]. In addition, the HIV integrase has been the focus of a number of smaller interaction screens [24–26]. Furthermore, host interactions of three other HIV proteins were identified with a mass spectrometry-based pull-down approach [27] and a small number of virus–host interactions for other viruses [28,29] (Table 1).

Whereas Y2H screens detect direct virus–host interactions, many other interactions are indirect and occur in protein complexes. These indirect interactions can only be detected by proteomic (pull-down assays and consecutive mass spectometry) or RNA interference (RNAi) screens. RNAi assays, however, do not only identify (i) host factors (HF) present in protein complexes, but also (ii) HFs involved in signaling pathways and cellular processes relevant for infection, as well as (iii) HFs binding to viral non-protein components (e.g. nucleic acids) [30]. In addition, genetic screens [31–33] and genome-wide association studies (GWAS) [34,35] provide hints on HFs involved in infection.

The first genome-wide RNAi screen of viral host factors was performed in 2005 by Cherry *et al.* for an insect picornavirus in a *Drosophila* system [36]. Several of the factors they identified were also crucial for virus proliferation of related mammalian viruses such as HCV. A similar strategy was later applied to dengue [37] and influenza virus [38]. In both cases, a genome-wide screen was performed in *Drosophila* cells and then homologs of identified HFs were validated in a targeted screen. The first large-scale RNAi screen of virus infection in human cells was performed by Ng *et al.* for HCV [39] with small interfering RNAs (siRNAs) against 4000 druggable genes. Several genome-scale screens were subsequently published involving HIV-1 [23,40–42], West Nile virus (WNV) [43], HCV [44,45] and influenza [46–48] (Table 2). Thus, a large number of studies on virus–host interactions, both direct and indirect, have been performed, providing considerable resources on these interactions for systems biology studies on virus infection.

## Reproducibility of large-scale virus–host interaction screens

In general, reproducibility of independent large-scale screens is very low even if the same technology is used. For example, in the two large-scale intracellular Y2H screens of human interactions by Stelzl *et al.* [49] and Rual *et al.* [50], only 19 interactions were detected in both screens. Similarly, in the recently published intraviral Y2H screen of five herpesvirus species [51], only six of 43 (13.9%) EBV interactions identified by Calderwood *et al.* [19] were recovered. One of the major reasons for the small overlap between different Y2H screens is the rather low sensitivity of this technique. A comparative analysis using positive and negative interaction reference sets determined a coverage of only approximately 20–30%, which is in the same range as all other assays for detection of binary interactions that were tested [52].

Several independent RNAi screens of HFs were performed for HIV-1, HCV and influenza virus [23,40,42,44–48]. For HIV-1, overlaps were very modest ranging from 3 to 6% [27] (Figure 1a). The largest overlaps were observed between the studies of Brass *et al.* [40] and Zhou *et al.* [42], which used the same cell lines (HeLa) and both focused on the entire virus life cycle. By contrast, König *et al.* [23] analyzed only the processes subsequent to HIV-1 entry and used different cell lines (293T). For HCV, Li *et al.* [45] recovered 15 of 96 (16%) HCV HFs from a previous genome-wide study by Tai *et al.* [44]. Nevertheless, relative to the size of the primary hit list by *Li et al.* [45], this also corresponds only to an overlap of 3%. Similarly, the overlap between the four RNAi studies on influenza virus in hu-

**Table 2. Overview of large-scale RNAi screens of host factors**

| Virus | Cell line | Library | Number of host factors | Refs |
|---|---|---|---|---|
| *Drosophila* C virus (DCV) | *Drosophila* S2 | *Drosophila* double-stranded RNA (dsRNA, ~21 000 dsRNAs) | 112 | [36] |
| Hepatitis C virus (HCV) | Huh-7/EN5-3 | siRNA (4000 druggable human genes, four siRNAs per pool) | 9 | [39] |
| Influenza (FVG-R) | *Drosophila* DL1 | dsRNA (*Drosophila*, 13 071 genes) | 110 | [38] |
| Influenza (FVG-R, H1N1, H5N1) | HEK293 | siRNAs against human homologs of three *Drosophila* HFs | 3 | [38] |
| Insect dengue virus (DENV-2) | *Drosophila* S2 | dsRNA DRSC 2.0 library (*Drosophila*, 22 632 dsRNAs) | 116 | [37] |
| Human dengue virus | HuH-7 | siRNAs against human homologs of 82 *Drosophila* HFs | 42 | [37] |
| HIV-1 | TZM-bl | siRNA (human, 21 121 genes, four siRNAs per pool) | 273 | [40] |
| West Nile virus (WNV) | HeLa | siRNA (human, 21 121 genes, four siRNAs per pool) | 283 HSFs[a] + 22 HRFs[b] | [43] |
| HIV-1 | 293T | siRNA (human, ~20 000 genes, six siRNAs per gene, two siRNAs per pool) | 295 | [23] |
| HIV-1 | TZM-bl | siRNA (human, 19 709 genes, three siRNA per pool) | 232 | [42] |
| HCV | Huh7/Rep-Feo | siRNA (human, 21 094 genes, four siRNAs per pool) | 96 | [44] |
| HCV | Huh 7.5.1 | siRNA (human, 19 470 genes, four siRNA per pool) | 262 | [45] |
| HIV-1 | Jurkat | Short hairpin (shRNA, human, 54 509 transcripts, 3–5 shRNA per transcript) | 252 | [41] |
| Influenza (A/PR/8/34) | HBEC | siRNA (1745 genes selected based on Y2H interactions and transcriptional regulation) | 616 | [22] |
| Influenza (A/Puerto Rico/8/34) | U2OS | siRNA (human, 17 877 genes, four siRNAs per pool) | 250 | [48] |
| Influenza (WSN-Ren) | A549 | siRNA (human, >19 000 genes, six different libraries) | 295 | [46] |
| Influenza (A/WSN/33) | A549 | siRNA (human,17 000 annotated and 6000 predicted genes, 62 000 siRNAs) | 287 | [47] |

[a]Host susceptibility factors (HSF): facilitate virus infection.

[b]Host resistance factors (HRF): reduce virus infection.

man cells was low and ranged from approximately 1 to 12%, with the highest overlap between the studies by König *et al.* [46] and Karlas *et al.* [47]. The reasons for the large discrepancies are most likely differences in the experimental setup, such as the cell culture systems, virus isolates and siRNA pools that were used in these studies. Further factors that largely influence the outcome of RNAi screens are the criteria used for filtering false positives and identifying the final set of published interactions. Indeed, Bushman *et al.* showed for the König *et al.* HIV-1 [23] data that the application of different toxicity thresholds significantly influenced the final hit lists [27]. Accordingly, because all studies used quite different criteria, the application of consistent criteria might increase the overlaps in HFs. In summary, different RNAi studies can only be appropriately compared if they used similar (i) experimental setups and (ii) cutoff and selection criteria.

Unsurprisingly, the overlap between published Y2H and RNAi screens are similarly low. Because these screens were performed and analyzed by distinct groups and not

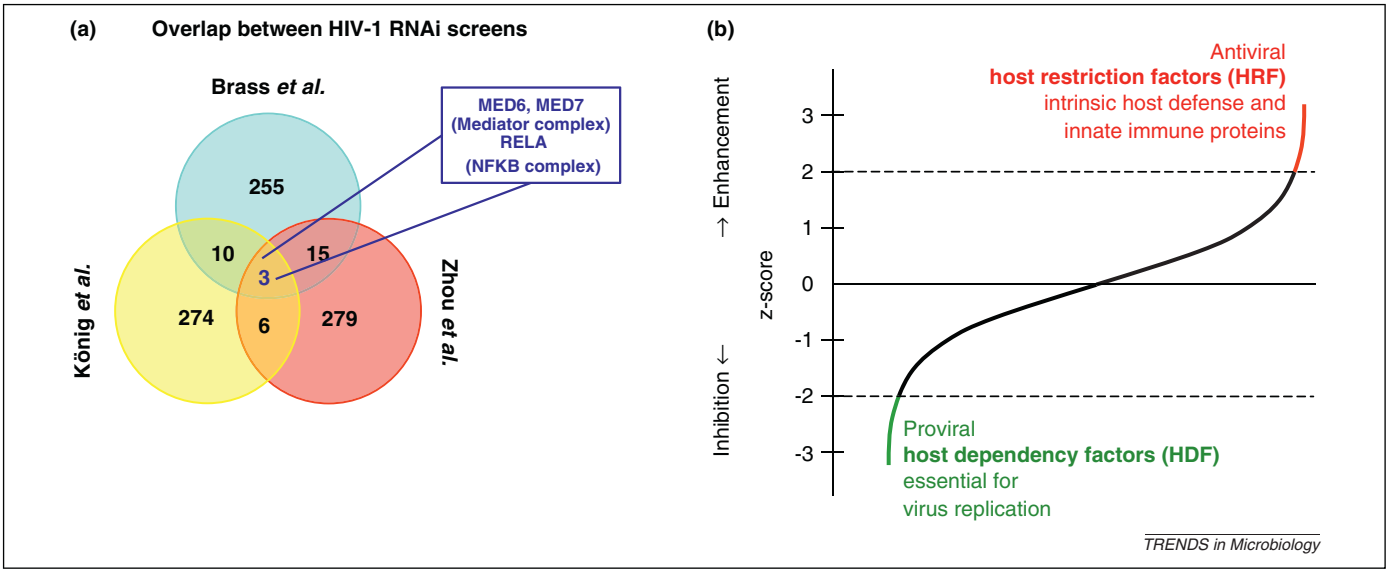


**Figure 1**. Overlap of large-scale screens and illustration of enrichment analysis. **(a)** Analysis of the HIV-1 host factors (HFs) identified in the three genome-scale siRNA screens by Brass *et al.* [40], König *et al.* [23] and Zhou *et al.* [42]. Numbers are taken from the meta-analysis by Bushman *et al.* [27]. Only three HFs were identified as hits by all three studies involving the Mediator and the NF-κB (NFKB) complexes. In pairwise comparisons, the overlap was between 3 and 6%. **(b)** To derive normalized RNAi results, Z-scores are calculated that indicate the number of standard deviations the observed value is above or below the mean. Genes are then ranked in the order of the Z-scores, from lowest (decreased infection upon gene knockdown, indicates host proteins essential for viral replication) to highest (increased infection after knockdown, indicates host defense and immune proteins).

under comparable conditions, these differences are not only due to the technological differences. Future studies will have to show whether a combined systematic analysis using both techniques will yield synergistic or rather complementary results, both of which would be helpful to elucidate critical events in viral infection.

## Cellular pathways and functions crucial to virus–host interaction

The genome-scale screens discussed above provide candidate genes important in some way for virus replication. However, to improve understanding of the viral life cycle, it is not sufficient to look only at individual genes, but information on the identified genes has to be integrated into a systems-level view of the processes involved. The most straightforward way to put the genes into the cellular context is to focus on the pathways and processes they are involved in, their functions and other properties. Here, several public databases provide such information. For example, pathway information can be obtained from the KEGG database [53] whereas the Gene Ontology (GO) database provides structured and standardized annotations of biological processes, molecular functions and cellular compartments for genes [54].

Although simply collecting all annotations for all genes on the hit list can give an idea of common properties of the identified genes, this can be misleading due to different background frequencies. For example, the number of nuclear genes is much larger than the number of subunits of the COPI coatomer complex. Accordingly, finding 10 subunits of the COPI complex among the targets would be highly interesting, whereas finding 10 nuclear proteins would not be. To identify functional categories that are truly overrepresented among the virus targets, an enrichment analysis is usually performed for which the overlap between hit list and functional category is determined and compared to the background frequency of the functional category. The most common approach is to calculate statistical significance for the enrichment, for instance using a hypergeometric test [55]. For this purpose, a range of alternative methods have been proposed [56] and several tools and web services have been developed such as the DAVID webserver [57]. Enrichment analysis was performed in several of the above mentioned studies but the results are difficult to compare due to the different methods and functional categories used. For example, although both the de Chassey *et al.* [20] and Shapira *et al.* [22] study identified WNT and MAPK signaling pathways as being targeted by HCV and H1N1, respectively, the definition of these pathways differed between the studies and Shapira *et al.* also included first neighbors of the direct host target proteins in the analysis. These differences make it difficult to assess similarities and differences in the enriched categories between different studies. An alternative approach was applied by König *et al.* in their studies of HIV-1 and influenza HFs [23,46]. Here, they used the so-called ontology-based pattern identification (OPI) algorithm [58], which identifies clusters of genes that show both a similar pattern in inhibiting or enhancing virus growth but also a statistically significant enrichment in some functional category.

To allow the comparison of targeted pathways and processes between different screens, Bushman *et al.* repeated enrichment analysis in their meta-analysis of HIV siRNA screens using the same approach for each screen [27]. They identified a number of functional categories enriched in all screens, such as the Mediator complex. Two of the Mediator subunits were detected in all three studies (Figure 1a). Interestingly, the overlap in enriched functional categories between the screens was greater than the overlap in the individual hit lists. This suggests that despite low sensitivity and specificity, high-throughput screens can still be successful in identifying the pathways and processes targeted by the respective virus. Similarly, in the meta study of published pathogen interacting proteins, Dyer *et al.* [6] found functional categories enriched for direct targets of several viruses and bacteria. Accordingly, such meta-analyses can provide a better understanding of the common pathways and processes targeted by pathogens.

## Systematic profiling of innate immune proteins in virus infection by RNAi

RNAi perturbation screens do not only identify HFs mandatory for infection, but also antiviral HFs that inhibit virus replication. Knockdown of these antiviral HFs in RNAi screens leads to increased infection rates. In fact, in normalized RNAi screening results, the number of inhibitory HFs that represent proteins involved in intrinsic host defense and innate immunity is about as high as the number of HFs essential for virus replication (Figure 1b). A considerable number of cellular proteins thus appear to restrict virus replication and might be considered as intrinsic host defense proteins. Thus, the current concept of just a few limiting host restriction factors might have to be redefined. However, depletion of a considerable number of innate immune proteins (involved for example in interferon induction or signaling) does not necessarily result in increased infection rates, which is probably caused by redundancy effects. In fact, depletion of some innate immune proteins, e.g. some defensin family members in the case of WNV [43], can even decrease infection rates, suggesting that these proteins possess a more complex role and actually inhibit the immune response to this particular pathogen. Accordingly, a detailed analysis of immunoregulatory HFs might help to refine the complex relationship between different immune signaling pathways.

## Centrality of viral targets and host factors

To investigate the importance of the identified viral interactors for the cellular host, the most common approach is to analyze the centrality of these proteins within the host interaction network. Here, centrality of a protein can be quantified in different ways such as the number of its interaction partners (degree) or the distances to other proteins in the network (Figure 2a). The distance between two proteins is defined as the length of the shortest path between them. Direct interaction partners have a distance of 1, a distance of 2 means that they are connected by interactions to a third protein, and so on. Both centrality measures were investigated by Calderwood *et al.* for the EBV–human interactions [19]. They found that virus proteins tend to interact

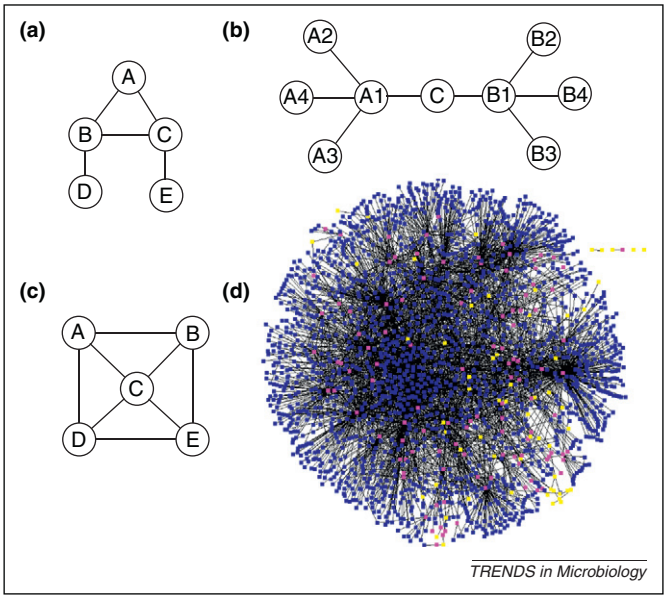


**Figure 2.** Network centrality measures. **(a)** The degree of a protein is defined as the number of its interactions. The degrees in the example are: A,B,C:2; D,E:1. The shortest path between any pair of proteins is defined as the smallest number of interactions that have to be followed to get from one protein to the other. In the example, proteins D and E are connected by two paths: D→B→A→C→E and D→B→C→E. Because the second path uses one less interaction, the shortest path length between D and E is 3. **(b)** Betweenness centrality of a protein C is defined by the fraction of shortest paths between any pair of proteins that pass protein C. In the example, every shortest path from any of the A1, A2, A3, A4 proteins to any of the B1, B2, B3, B4 proteins has to pass C. Accordingly, betweenness centrality of C is high despite a small degree and C can be considered a bottleneck of the network. This example illustrates that betweenness centrality and degree are not necessarily correlated. Another example is shown in **(c)**. Protein C has the highest degree in the network (4) but a low betweenness centrality. Because most other proteins are directly connected, only a small fraction of shortest paths between any pair of these proteins crosses C. **(d)** High degree centrality of host proteins that interact with EBV proteins. Virus–host and intraviral interactions are taken from the study of Calderwood *et al.* [19]. Viral proteins are shown in yellow, direct interactors of the viral proteins in pink and proteins interacting with the direct interactors in dark blue.

with central proteins in the host interaction network that are both highly connected (hubs) and have small distances to the remaining network (Figure 2b,d). These results were confirmed by Dyer *et al.* [6] for both viruses and bacteria and a preference was found for viral proteins to interact with so-called bottlenecks, which are defined as proteins with high betweenness centrality. Betweenness centrality of a protein $P$ is determined as the fraction of shortest paths between any pair of proteins passing through $P$, summed over all possible pairs of proteins (Figure 2b,c). Although degree and betweenness centrality are correlated, the enrichment of bottlenecks among the viral targets was found to be independent of their high degree [6,20].

Similar results were also reported by for direct interactors of HCV [20] and HIV [59], as well as HFs of influenza [46] and HIV-1 [59,60]. Furthermore, the subnetworks between viral interaction partners or HFs have been found to be densely connected with many more interactions between them than expected at random [19,20,22,59]. There is evidence that essentiality of a protein for cell survival is correlated both to the degree and betweenness centrality of proteins in the cellular network [61–64]. Thus, the abundance of central proteins among virus targets and HFs suggests that viruses tend to interact with and depend on essential host proteins. Furthermore, the number of pathways a host protein is involved in increases with its degree [59]. Thus, by targeting a relatively small number of hubs, viruses can interact with a large number of cellular pathways amplifying their effect on the cellular system. Indeed, for HIV-1 direct interactors and HFs, the number of pathways they interact with is increased significantly compared to the background of proteins in the cellular network [59]. The biological implications of these results are currently not fully understood, as well as whether there are differences that relate to distinct features of these pathogens, for example distinct replication strategies of DNA versus RNA viruses.

## Finding functional modules vital to the virus life cycle

Because enrichment analysis searches only for pre-defined functional categories overrepresented among the viral targets and HFs, network-based approaches have been developed to identify novel modules of functionally-related genes among these genes. The first to pursue such an approach for a large-scale screen were König *et al.* [23] for HIV-1. They applied a network clustering algorithm (MCODE) [65] to the cellular interaction network between the identified HFs. MCODE identifies clusters of proteins within the network with high local density, i.e. many interactions between them (Figure 3a). Here, network clustering was used to select the final hit list and to identify interesting subnetworks with high connectivity. Although they could identify several dense subnetworks, these mainly reflected the functional categories already identified using enrichment analysis and the OPI algorithm.

By contrast, Bushman *et al.* used MCODE to detect several functional modules in the cellular network of HIV-1 HFs from all genome-scale studies and additional interacting proteins that were not found in their enrichment analysis [27]. Interestingly, the module with the highest number of proteins identified in two or more screens contains several subunits of the Mediator complex, which was also significant in the enrichment analysis. For the other clusters, proteins were generally only found in one or few screens suggesting that the corresponding functional categories would only show up in an enrichment analysis on all screens combined. The disadvantage of this kind of analysis is that, although it does not depend on predefined categories, significance of the identified modules is difficult to assess. Furthermore, clusters are difficult to interpret if they do not correspond to previously described functional categories or pathways. For instance, although König *et al.* [46] identified 152 clusters in their siRNA screen on influenza HFs, only 77 of these were biologically interpretable because they were enriched in a functional category.

An alternative approach to finding interesting clusters among virus interacting proteins was presented by MacPherson *et al.* [66]. Instead of clustering the target genes based on their interactions within the host cell, they used a biclustering approach to identify HIV interacting proteins sharing a common set of HIV interactions (Figure 3b). Biclustering was used to cluster the interactions in addition to the genes. Because genes can then be contained in several clusters, a bottom-up tree-building algorithm was used to recover the relationships between clusters. By partitioning this tree, they defined 37 subsystems enriched

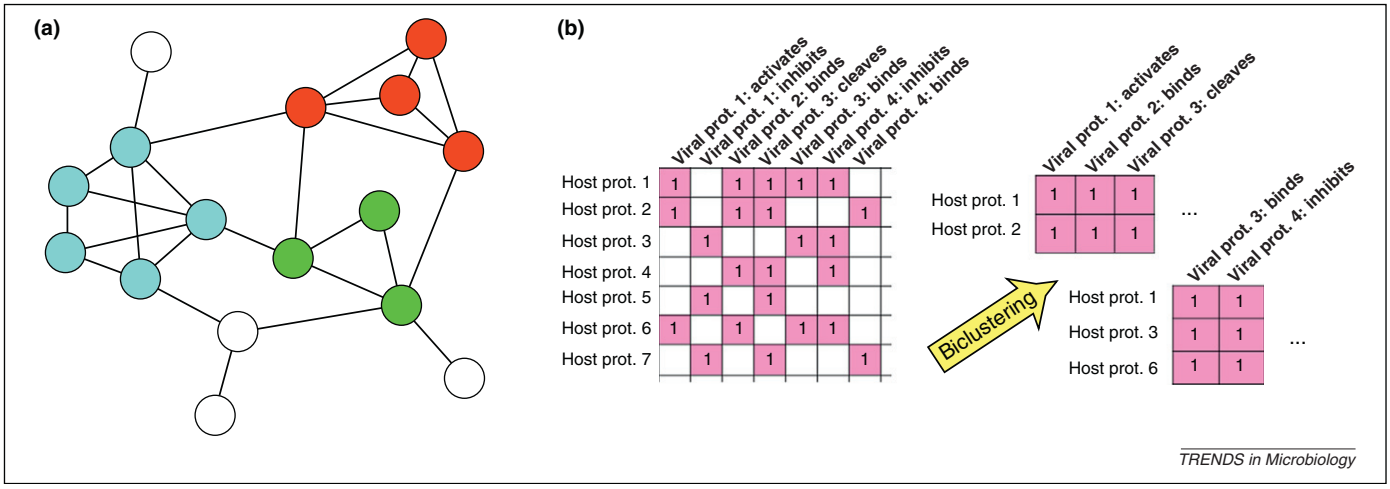


**Figure 3**. Identification of functional modules in the virus–host interaction network. This figure illustrates two alternative approaches to identify novel functional modules in the virus–host interactomes. **(a)** Illustration of network clustering approaches that try to determine densely connected modules between hits identified in large-scale screens. For this purpose, hits are mapped to the cellular interaction network of the host. Density of subnetworks is commonly defined in terms of the number of interactions between the proteins in the subnetworks compared to the number of possible interactions. The higher the ratio between the two values, the higher the density. In the example, three densely connected subnetworks can be identified, shown in red, green and blue. **(b)** Biclustering of proteins (prot.) and interactions at the same time (illustration based on MacPherson *et al.* [66]). Biclustering is applied to a matrix with the host proteins as rows and the combination of virus proteins and interaction type as columns. The algorithm then tries to find subsets of host proteins which have several interactions in common. Here, each host protein may be contained in different subsets characterized by different interactions. In this example, host protein 1, for example, is contained both in a bicluster with host protein 2 characterized by three common interactions and a bicluster with host proteins 3 and 6, which is characterized by a different set of two interactions.

for specific functional categories. Unfortunately, this approach relies heavily on the availability of information specifying different types of interaction in addition to the viral interaction partner, which limits its applicability to large-scale Y2H or RNAi screens. Nevertheless, MacPherson *et al.* used the identified clusters to interpret results from the three HIV siRNA screens. Of the 37 subsystems, 10 groups were significantly enriched in either an individual siRNA screen or all combined. However, the five most significant subsystems were already identified by Bushman *et al.* [27] by a simple enrichment analysis. It remains to be seen how relevant the remaining five novel categories are because they were either only significant in one of the HF lists or only weakly significant in the combination of all lists.

These results show that network clustering approaches can yield new interesting insights into novel functional modules important for viral replication which a simple enrichment analysis on known functional categories might fail to identify.

## Concluding remarks

Advances in high-throughput screening technologies have opened up new avenues for studying global aspects of virus infection and identifying potential host target proteins for antiviral therapy. Although hundreds of potential target proteins were identified in several Y2H and RNAi screens, overlaps between independent studies for the same virus have been modest at best even if the same technology was used. Although standardization of experimental protocols and hit calling strategies could increase consistency between independent screens, a large fraction of discrepancies are due to the intrinsic low sensitivity and specificity of large-scale methods. Furthermore, because different host proteins could be required for virus proliferation under different conditions, repetition of screens in varying cell lines might provide a more complete picture of the processes involved. Because Y2H and RNAi screens identify different types of interactions between the virus and the host cell, parallel screens with both techniques might provide complementary information.

To interpret the identified viral interactors and put them in the cellular context, computational approaches are crucial. The most common approach is to determine functions and processes enriched among the virus targets. Although this approach is dependent on predefined categories, more sophisticated approaches, e.g. based on network clustering, are not necessarily more successful in identifying interesting pathways or functions targeted by the virus. Although they may identify novel concepts, significance and biological relevance is often difficult to assess without referring to predefined functional annotations. Additional focus has been put on the high centrality of the host factors within the cellular protein interaction networks. This suggests that these proteins are important for many different pathways and processes in the cell and has two implications for the development of antiviral drugs. First, it suggests that more central host proteins are also more important for virus infection and, thus, interesting drug targets. Second, drugs against these proteins will affect many pathways and processes likely resulting in considerable adverse effects. To date, no concrete drug targets have been developed based on large-scale screens. However, an overlap of HFs in HIV with druggable protein domains identified at least a few potential targets, some of which are known to be targeted by small molecules with anti-HIV activity [27].

In summary, application of both experimental and computational systems biology approaches can lead to crucial insights into virus infection and its life cycle. However, there is still considerable need for improvement in the precision of screening methods and approaches for subsequent analysis. To gain improved global models of virus-infected cells that include spatial and temporal in-

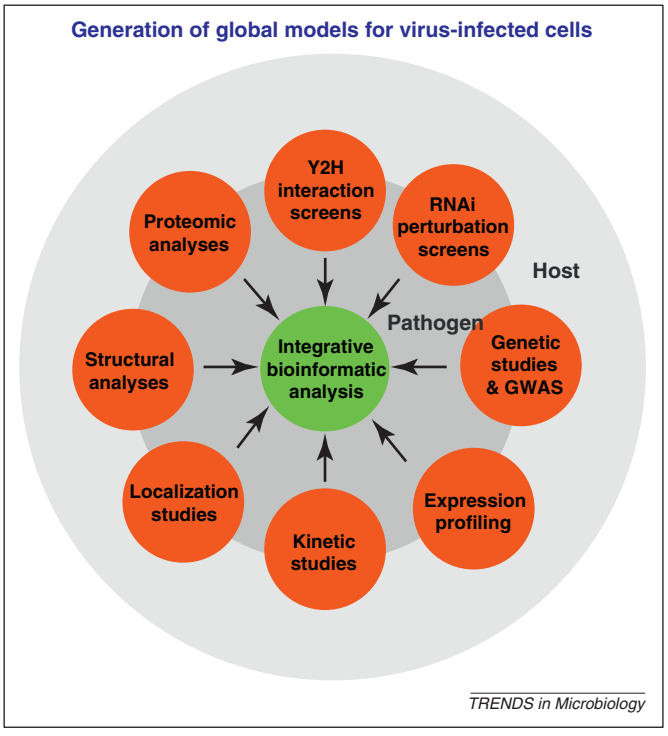


**Figure 4**. Analysis of virus–host interactomes by integration of different technologies. Development of integrative computational approaches that combine results of different experimental technologies, as outlined in this figure, will be of crucial importance for the analysis and understanding of virus–host interactomes.

formation, approaches have to be developed that integrate different sources of information, such as data from technologies that detect direct and indirect protein interactions, expression profiling, localization and infection kinetics (Figure 4). Genome-scale localization studies, for example, enable functional predictions for viral proteins [67], and, in combination with protein interaction data, provide insight into possible regulatory mechanisms [68]. The integrated analysis of array- or deep sequencing-based GWAS studies with genome-scale protein interaction and transcriptional profiling screens might allow identifying polymorphic host factors that crucially influence disease and treatment outcome. Biomarkers identified by such approaches will very probably contribute significantly to future changes in healthcare towards personalized medicine. Alternative strategies include the combined application of several different technologies in one experimental approach (e.g. transcriptional profiling in siRNA knockdown cells) [69], and the application of iterative or flexible-design screens. In such integrative approaches, we believe, lies the important contribution of systems biology to the understanding and eventual treatment of viral diseases.

## References

1 UN Joint Programme on HIV/AIDS (2010) *UNAIDS Report on the Global AIDS Epidemic: 2010*, http://www.unaids.org/globalreport/documents/20101123_GlobalReport_full_en.pdf.
2 Johnson, N.P. and Mueller, J. (2002) Updating the accounts: global mortality of the 1918-1920 'Spanish' influenza pandemic. *Bull. Hist. Med.* 76, 105–115
3 Rickinson, A. (2002) Epstein-Barr virus. *Virus Res.* 82, 109–113
4 Rezk, S.A. and Weiss, L.M. (2007) Epstein-Barr virus-associated lymphoproliferative disorders. *Hum. Pathol.* 38, 1293–1304
5 Martin, D. and Gutkind, J.S. (2008) Human tumor-associated viruses and new insights into the molecular mechanisms of cancer. *Oncogene* 27 (Suppl. 2), S31–S42
6 Dyer, M.D. *et al.* (2008) The landscape of human proteins interacting with viruses and other pathogens. *PLoS Pathog.* 4, e32
7 Ptak, R.G. *et al.* (2008) Cataloguing the HIV type 1 human protein interaction network. *AIDS Res. Hum. Retroviruses* 24, 1497–1502
8 Fu, W. *et al.* (2009) Human immunodeficiency virus type 1, human protein interaction database at NCBI. *Nucleic Acids Res.* 37, D417–D422
9 Navratil, V. *et al.* (2009) VirHostNet: a knowledge base for the management and the analysis of proteome-wide virus–host interaction networks. *Nucleic Acids Res.* 37, D661–D668
10 Chatr-aryamontri, A. *et al.* (2009) VirusMINT: a viral protein interaction database. *Nucleic Acids Res.* 37, D669–D673
11 Driscoll, T. *et al.* (2009) PIG: the pathogen interaction gateway. *Nucleic Acids Res.* 37, D647–D650
12 Kumar, R. and Nanduri, B. (2010) HPIDB: a unified resource for host–pathogen interactions. *BMC Bioinform.* 11 (Suppl. 6), S16
13 McCraith, S. *et al.* (2000) Genome-wide analysis of vaccinia virus protein–protein interactions. *Proc. Natl. Acad. Sci. U.S.A.* 97, 4879–4884
14 Bartel, P.L. *et al.* (1996) A protein linkage map of *Escherichia coli* bacteriophage T7. *Nat. Genet.* 12, 72–77
15 von Brunn, A. *et al.* (2007) Analysis of intraviral protein–protein interactions of the SARS coronavirus ORFeome. *PLoS ONE* 2, e459
16 Flajolet, M. *et al.* (2000) A genomic approach of the hepatitis C virus generates a protein interaction map. *Gene* 242, 369–379
17 Choi, I.R. *et al.* (2000) Multiple interactions among proteins encoded by the mite-transmitted wheat streak mosaic tritimovirus. *Virology* 267, 185–198
18 Guo, D. *et al.* (2001) Towards a protein interaction map of potyviruses: protein interaction matrixes of two potyviruses based on the yeast two-hybrid system. *J. Gen. Virol.* 82, 935–939
19 Calderwood, M.A. *et al.* (2007) Epstein-Barr virus and virus human protein interaction maps. *Proc. Natl. Acad. Sci. U.S.A.* 104, 7606–7611
20 de Chassey, B. *et al.* (2008) Hepatitis C virus infection protein network. *Mol. Syst. Biol.* 4, 230
21 Zhang, L. *et al.* (2009) Analysis of vaccinia virus–host protein–protein interactions: validations of yeast two-hybrid screenings. *J. Proteome Res.* 8, 4311–4318
22 Shapira, S.D. *et al.* (2009) A physical and regulatory map of host–influenza interactions reveals pathways in H1N1 infection. *Cell* 139, 1255–1267
23 König, R. *et al.* (2008) Global analysis of host–pathogen interactions that regulate early-stage HIV-1 replication. *Cell* 135, 49–60
24 Studamire, B. and Goff, S.P. (2008) Host proteins interacting with the Moloney murine leukemia virus integrase: multiple transcriptional regulators and chromatin binding factors. *Retrovirology* 5, 48
25 Rain, J.C. *et al.* (2009) Yeast two-hybrid detection of integrase-host factor interactions. *Methods* 47, 291–297
26 Allouch, A. and Cereseto, A. (2009) Identification of cellular factors binding to acetylated HIV-1 integrase. *Amino Acids* DOI: 10.1007/s00726-009-0444-3
27 Bushman, F.D. *et al.* (2009) Host cell factors in HIV replication: meta-analysis of genome-wide studies. *PLoS Pathog.* 5, e1000437
28 Moerdyk-Schauwecker, M. *et al.* (2011) Detecting protein–protein interactions in vesicular stomatitis virus using a cytoplasmic yeast two hybrid system. *J. Virol. Methods* 173, 203–212
29 Covaleda, L. *et al.* (2010) Identification of cellular proteins interacting with equine infectious anemia virus S2 protein. *Virus Res.* 151, 235–239
30 Cherry, S. (2009) What have RNAi screens taught us about viral–host interactions? *Curr. Opin. Microbiol.* 12, 446–452
31 Yu, F. *et al.* (2007) Systematic identification of cellular signals reactivating Kaposi sarcoma-associated herpesvirus. *PLoS Pathog.* 3, e44
32 Sui, B. *et al.* (2009) The use of Random Homozygous Gene Perturbation to identify novel host-oriented targets for influenza. *Virology* 387, 473–481
33 Carette, J.E. *et al.* (2009) Haploid genetic screens in human cells identify host factors used by pathogens. *Science* 326, 1231–1235
34 Ge, D. *et al.* (2009) Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. *Nature* 461, 399–401

35 Puthothu, B. *et al.* (2006) TLR-4 and CD14 polymorphisms in respiratory syncytial virus associated disease. *Dis. Markers* 22, 303–308
36 Cherry, S. *et al.* (2005) Genome-wide RNAi screen reveals a specific sensitivity of IRES-containing RNA viruses to host translation inhibition. *Genes Dev.* 19, 445–452
37 Sessions, O.M. *et al.* (2009) Discovery of insect and human dengue virus host factors. *Nature* 458, 1047–1050
38 Hao, L. *et al.* (2008) Drosophila RNAi screen identifies host genes important for influenza virus replication. *Nature* 454, 890–893
39 Ng, T.I. *et al.* (2007) Identification of host genes involved in hepatitis C virus replication by small interfering RNA technology. *Hepatology* 45, 1413–1421
40 Brass, A.L. *et al.* (2008) Identification of host proteins required for HIV infection through a functional genomic screen. *Science* 319, 921–926
41 Yeung, M.L. *et al.* (2009) A genome-wide short hairpin RNA screening of jurkat T-cells for human proteins contributing to productive HIV-1 replication. *J. Biol. Chem.* 284, 19463–19473
42 Zhou, H. *et al.* (2008) Genome-scale RNAi screen for host factors required for HIV replication. *Cell Host Microbe* 4, 495–504
43 Krishnan, M.N. *et al.* (2008) RNA interference screen for human genes associated with West Nile virus infection. *Nature* 455, 242–245
44 Tai, A.W. *et al.* (2009) A functional genomic screen identifies cellular cofactors of hepatitis C virus replication. *Cell Host Microbe* 5, 298–307
45 Li, Q. *et al.* (2009) A genome-wide genetic screen for host factors required for hepatitis C virus propagation. *Proc. Natl. Acad. Sci. U.S.A.* 106, 16410–16415
46 König, R. *et al.* (2010) Human host factors required for influenza virus replication. *Nature* 463, 813–817
47 Karlas, A. *et al.* (2010) Genome-wide RNAi screen identifies human host factors crucial for influenza virus replication. *Nature* 463, 818–822
48 Brass, A.L. *et al.* (2009) The IFITM proteins mediate cellular resistance to influenza A H1N1 virus, West Nile virus, and dengue virus. *Cell* 139, 1243–1254
49 Stelzl, U. *et al.* (2005) A human protein–protein interaction network: a resource for annotating the proteome. *Cell* 122, 957–968
50 Rual, J.F. *et al.* (2005) Towards a proteome-scale map of the human protein–protein interaction network. *Nature* 437, 1173–1178
51 Fossum, E. *et al.* (2009) Evolutionarily conserved herpesviral protein interaction networks. *PLoS Pathog.* 5, e1000570
52 Braun, P. *et al.* (2009) An experimentally derived confidence score for binary protein–protein interactions. *Nat. Methods* 6, 91–97
53 Kanehisa, M. *et al.* (2010) KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.* 38, D355–D360
54 Ashburner, M. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25, 25–29
55 Khatri, P. and Draghici, S. (2005) Ontological analysis of gene expression data: current tools, limitations, and open problems. *Bioinformatics* 21, 3587–3595
56 Huang da, W. *et al.* (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37, 1–13
57 Huang da, W. *et al.* (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57
58 Zhou, Y. *et al.* (2005) In silico gene function prediction using ontology-based pattern identification. *Bioinformatics* 21, 1237–1245
59 Wuchty, S. *et al.* (2010) Viral organization of human proteins. *PLoS ONE* 5, e11796
60 van Dijk, D. *et al.* (2010) Identifying potential survival strategies of HIV-1 through virus–host protein interaction networks. *BMC Syst. Biol.* 4, 96
61 Hahn, M.W. and Kern, A.D. (2005) Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. *Mol. Biol. Evol.* 22, 803–806
62 Yu, H. *et al.* (2004) Genomic analysis of essentiality within protein networks. *Trends Genet.* 20, 227–231
63 Yu, H. *et al.* (2007) The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. *PLoS Comput. Biol.* 3, e59
64 Jeong, H. *et al.* (2001) Lethality and centrality in protein networks. *Nature* 411, 41–42
65 Bader, G.D. and Hogue, C.W. (2003) An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinform.* 4, 2
66 MacPherson, J.I. *et al.* (2010) Patterns of HIV-1 protein interaction identify perturbed host-cellular subsystems. *PLoS Comput. Biol.* 6, e1000863
67 Salsman, J. *et al.* (2008) Genome-wide screen of three herpesviruses for protein subcellular localization and alteration of PML nuclear bodies. *PLoS Pathog.* 4, e1000100
68 Sander, G. *et al.* (2008) Intracellular localization map of human herpesvirus 8 proteins. *J. Virol.* 82, 1908–1922
69 Amit, I. *et al.* (2009) Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses. *Science* 326, 257–263
70 Randall, G. *et al.* (2007) Cellular cofactors affecting hepatitis C virus infection and replication. *Proc. Natl. Acad. Sci. U.S.A.* 104, 12884–12889