

RESEARCH

Open Access

An integrative approach for a network based meta-analysis of viral RNAi screens

Sandeep S Amberkar^{1,2,3} and Lars Kaderali^{1,2*}

Abstract

Background: Big data is becoming ubiquitous in biology, and poses significant challenges in data analysis and interpretation. RNAi screening has become a workhorse of functional genomics, and has been applied, for example, to identify host factors involved in infection for a panel of different viruses. However, the analysis of data resulting from such screens is difficult, with often low overlap between hit lists, even when comparing screens targeting the same virus. This makes it a major challenge to select interesting candidates for further detailed, mechanistic experimental characterization.

Results: To address this problem we propose an integrative bioinformatics pipeline that allows for a network based meta-analysis of viral high-throughput RNAi screens. Initially, we collate a human protein interaction network from various public repositories, which is then subjected to unsupervised clustering to determine functional modules. Modules that are significantly enriched with host dependency factors (HDFs) and/or host restriction factors (HRFs) are then filtered based on network topology and semantic similarity measures. Modules passing all these criteria are finally interpreted for their biological significance using enrichment analysis, and interesting candidate genes can be selected from the modules.

Conclusions: We apply our approach to seven screens targeting three different viruses, and compare results with other published meta-analyses of viral RNAi screens. We recover key hit genes, and identify additional candidates from the screens. While we demonstrate the application of the approach using viral RNAi data, the method is generally applicable to identify underlying mechanisms from hit lists derived from high-throughput experimental data, and to select a small number of most promising genes for further mechanistic studies.

Keywords: Network analysis, RNAi screening, Virus-host interactions

Background

RNA interference (RNAi) has become an important workhorse of functional genomics, and genome-wide RNAi screens have been employed for example to identify genes involved in cell growth and viability, proliferation, differentiation, signaling or trafficking [1-9]. The technology has furthermore accelerated the discovery of novel host dependency factors (HDF) and host restriction factors (HRF) in viral infection [10-19]. However, while RNAi is a very powerful tool to identify genes involved in a

specific biological process, the placement of hits in their functional and spatiotemporal context in the underlying molecular processes remains a major challenge [20,21]. The interpretation of RNAi data in particular for virus screens is complicated further by the observed low overlap between identified host factors, even in different screens targeting the same virus [22-24]. This low overlap has been explained by different experimental conditions such as host cell type and viral strain used, transfection, incubation and infection time, and siRNA library used [24] as well as by technical artifacts arising from cell population context [25,26]. Furthermore, due to the typical setup of RNAi experiments with primary screens followed by secondary validation assays, it is likely that published hit lists are highly specific, but not very sensitive, further

*Correspondence: lars.kaderali@tu-dresden.de

¹ Institute of Medical Informatics and Biometry, Medical Faculty, TU Dresden, Fetscherstr. 74, 01307 Dresden, Germany

² ViroQuant Research Group Modeling, BioQuant, Heidelberg University, INF 267, 69120 Heidelberg, Germany

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

explaining the low overlap observed between different screens at the level of individual genes [27]. This, however, severely restricts a comparative analysis of inter-species RNAi screens [28]. On the other hand, protein interaction networks, virus-host interaction networks and other heterogeneous data have increased tremendously [29-34]. This offers novel ways to interpret hit lists from RNAi experiments from a network perspective, by integrating individual hits in their systemic context. It has been shown that this approach increases the overlap between different screens for the same virus at the pathway level [24], and the method can be extended to meta-analysis of screens targeting different viruses. Being less dependent on individual genes, but rather focusing on pathways, may shed new light onto virus-specific and generic host processes facilitating or restricting infection, and may prove a promising approach to identify potential host targets for antiviral drug development.

Several meta-analyses of RNAi screens have been conducted, albeit most work focused on integrating different screens targeting a single virus [24,28,35,36]. A notable exception is the study by Snijder et al., including 45 screens targeting 17 different mammalian viruses [37]. The authors show that accounting for cellular heterogeneity improves gene overlaps between screens, but the study does not focus on functional regions within the host protein network targeted by different viruses. In contrast, Navratil et al. study virus-host protein interactions in the human interferon network [32], throwing light on how viruses of different families target the innate immune system. Other similar analyses focused largely on HIV, for example, Murali et al. employed a semi-supervised machine learning approach mapping RNAi hits onto a protein interaction network to predict new HDFs [38]. Macpherson et al. and similarly Maulik et al. mine the HIV-1 human protein interaction network using biclustering, and identify biclusters enriched with GO terms and RNAi hits [39,40]. Several authors have furthermore used protein-protein interaction (PPI) networks to identify topological properties of proteins targeted by pathogens. Dyer et al. characterized host proteins targeted by 190 different pathogens, including 35 viruses, 17 bacterial and two protozoan groups [29]. One of the major outcomes of this analysis was that pathogens preferentially target proteins with high node betweenness (bottlenecks) or high degree (hubs). Similarly, the studies by Dijk et al. and Dickerson et al. both showed that HIV preferentially targets hub and bottleneck genes in the human protein network [30,31]. Further characterizing the neighborhood of HDFs, Gulbahce et al. showed that proteins translated from genes involved in viral diseases are most likely located in the neighborhood of their corresponding viral targets [33].

Given the typically low overlap between different RNAi screens at the gene level and the relatively long hit lists resulting from individual screens, a central problem is how to select most promising candidates for functional characterization and detailed biochemical follow-up experiments. When looking for putative antiviral drug targets, one is typically interested in candidates that have a significant impact on infection outcome in the specific virus under consideration, or possibly even in several different viral species if e.g. broadly acting antivirals are sought for. Corresponding target pathways should therefore be “enriched” by hit genes from the RNAi data, while at the same time it is desirable that the respective targets are centrally located in the virus-host interaction network.

In this manuscript, we present a comparative analysis of RNAi hits for different viruses in the context of functional modules of protein interaction networks. The main purpose of our work is in hit prioritization, that is, we strive to identify a small set of candidates for further detailed follow-up experiments. We cluster the host protein network to identify functional host modules, and then use a statistical test to identify modules enriched with hits from seven genome-wide RNAi screens for three different viruses. Network topological characteristics are used to filter relevant subnetworks further, and resulting modules and their neighborhoods are annotated and interpreted. Using this approach, we identified several interesting candidate pathways for human immunodeficiency virus 1 (HIV-1) and hepatitis C virus (HCV), including known targets such as the mediator complex or members of the heterogeneous nuclear ribonucleoprotein subunits (hnRNPs) in HIV infection, or MAP kinases and heat shock proteins in HCV infection. Furthermore, using our approach, we predict that SERCA1 and Tankyrase-1 (TNKS1) may be interesting targets for further characterization in HCV infection.

Materials and methods

An overview of the data analysis pipeline used is shown in Figure 1. In brief, we collate information from 11 different public protein-protein interaction (PPI) data repositories, and integrate them into a large human PPI network. Subsequently, we use a cohesiveness-based greedy clustering algorithm to identify –possibly overlapping– clusters in the protein network, which are then tested for enrichment of hits from one or several RNAi screens. Significant modules are then filtered further using topological properties and semantic similarity, and functionally characterized using gene ontology and Reactome pathways. Using tissue-specific expression data, we predict novel putative host factors based on neighborhood relations in identified modules. We describe each of these steps in more detail in the following.

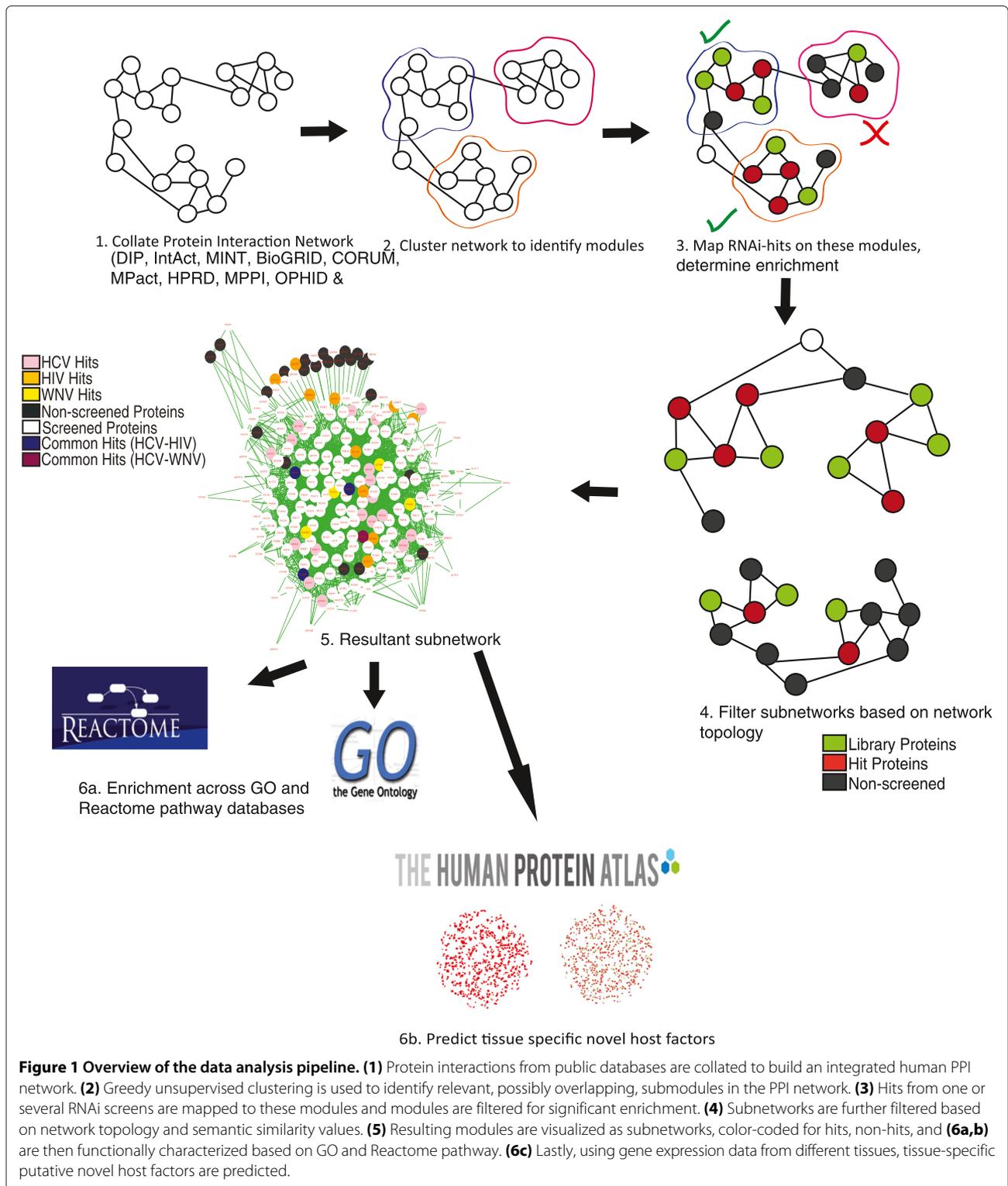


Figure 1 Overview of the data analysis pipeline. (1) Protein interactions from public databases are collated to build an integrated human PPI network. **(2)** Greedy unsupervised clustering is used to identify relevant, possibly overlapping, submodules in the PPI network. **(3)** Hits from one or several RNAi screens are mapped to these modules and modules are filtered for significant enrichment. **(4)** Subnetworks are further filtered based on network topology and semantic similarity values. **(5)** Resulting modules are visualized as subnetworks, color-coded for hits, non-hits, and **(6a,b)** are then functionally characterized based on GO and Reactome pathway. **(6c)** Lastly, using gene expression data from different tissues, tissue-specific putative novel host factors are predicted.

Human protein interaction network:

The human protein interaction network was collated from two major resources: the iRefIndex database, a meta-database comprising data from ten resources (DIP, IntAct,

MINT, BioGRID, BIND, CORUM, MPact, HPRD, MPPI, OPHID [41-51]), and the String v9.0 database [52] which includes both experimentally validated as well as computationally predicted interactions. The union of reported

interactions in these databases was used to establish our PPI network. We utilized a score filter of 0.75 on the STRING interactions as a tradeoff between reliability of included interactions and sufficient network density for further computations. Different thresholds between 0.6 and 0.9 were tested for the predicted interactions from STRING. For higher scores, the predicted interactions did not add much to the existing pool of interactions, and subsequent clustering resulted in few to no subnetworks. Conversely, for lower scores, the subnetworks included broad networks with multiple, non-specific functional annotations. A score of 0.75 led to optimal subnetworks that were functionally specific, and returned a reasonable number of subnetworks for further analysis. The overall procedure resulted in a protein interaction network comprising 15,383 proteins and 337,413 interactions from STRING and iRefIndex.

RNAi screening data:

We then mapped data from seven published genome-wide RNAi screens to the PPI network, including three human immunodeficiency virus-1 (HIV-1) screens [10,12,53], three Hepatitis C Virus (HCV) screens [13,18,54] and one west Nile virus (WNV) screen [11]. Further data analysis was then performed individually using only screens targeting the same virus (intra-species), as well as across all seven screens (inter-species).

Submodule identification and statistical testing:

We used the ClusterONE algorithm to detect overlapping subnetworks in the human PPI network. ClusterONE is a neighborhood-expansion, greedy graph clustering algorithm [55]. It is able to take edge weights corresponding to confidence scores into account in the clustering, and allows overlapping clusters where individual proteins may be part of more than one cluster. We used default values for most parameters of the ClusterONE algorithm, except for the `merge-method` parameter which was set to `multi` to merge highly overlapping clusters, as well as the minimum cluster size parameter, which we varied between 25 and 100. The variation of the cluster size parameter leads to clusters of different granularity, from very small, highly cohesive clusters, to larger and more heterogeneous clusters. Both may be desirable for the analysis of virus-targeted subnetworks, we therefore continued analysis with a redundant set of larger and smaller, overlapping clusters; we label this set of clusters \mathcal{C}^{all} in the following. Note that these clusters are not merged or integrated further, but rather \mathcal{C}^{all} is a set of different clusters. After clustering, we tested for significant enrichment of RNAi hits within each cluster in \mathcal{C}^{all} using Fisher's exact test, with significance level $\alpha = 0.05$, resulting in the set $\mathcal{C}^{hit} \subset \mathcal{C}^{all}$ of clusters significantly enriched with RNAi hits. We note that the clusters in \mathcal{C}^{hit} may still overlap and

may even contain clusters that are subsets/supersets of one another.

Submodule filtering and cluster selection:

We next used additional filtering criteria to select a small number of relevant clusters from \mathcal{C}^{hit} for further manual analysis. The underlying idea is to choose clusters that differ significantly from non-significant clusters not only based on their enrichment with RNAi hits, but also with respect to their "importance" in the underlying host PPI network. We selected seven network centrality measures and two further similarity measures for this filtering step. We briefly review these measures in the following, but before repeat some elementary definitions from graph theory.

Let $G = (V, E)$ be an undirected graph with nodes $v \in V$ corresponding to proteins and undirected edges $e \in E$ corresponding to interactions between proteins. As we consider undirected edges only, let $e_{ij} = e_{ji}$. We define a *path* P between two nodes $s, t \in V$ in a graph $G = (V, E)$ as a sequence $v_0, e_0, v_1, e_1, \dots, v_{k-1}, e_{k-1}, v_k$ of nodes $v_i \in V$ and edges $e_i \in E$, where edge e_i connects nodes v_i and v_{i+1} , where $v_i \neq v_j$ for all nodes in P , and where $v_0 := s$ and $v_k := t$. The length of P is defined as the number of edges in the path P .

When clustering the graph G using a graph clustering algorithm such as ClusterONE, the nodes V in G are grouped into different clusters. Let $V_C \subseteq V$ be one such cluster. This cluster induces a subnetwork $S_C = (V_C, E_C)$ on G , where $E_C = \{e_{ij} \in E : v_i, v_j \in V_C\}$, i.e., the induced subnetwork consists of the subset V_C of nodes, and all edges in E between these nodes in the original graph G . Hereafter, we use the term *subnetwork* to denote the full subnetwork $S_C = (V_C, E_C)$, whereas by *cluster* we refer only to the subset of nodes $V_C \subseteq V$.

To filter significant clusters $V_C \in \mathcal{C}^{hit}$ further, we used the following topological properties of the nodes in V_C respectively their induced subnetwork S_C :

1. **Average node degree:** The *node degree* of a vertex v in a graph $G = (V, E)$ is given by

$$\deg(v, G) := |\{e_{v,w} \in E \mid \forall w \in V\}|,$$

i.e., it is the number of edges in E adjacent to v . The *average node degree* of a subnetwork $S_C = (V_C, E_C)$ of G is the average degree of all nodes in V_C :

$$C_D(S_C) = \frac{1}{|V_C|} \sum_{v \in V_C} \deg(v, S_C),$$

where $|V_C|$ denotes the number of nodes in V_C . Note that we compute the degree with respect to the edge set E_C of the subgraph S_C , and not the full graph G .

2. **Average node betweenness:** The *node betweenness* of a node $v \in V$ is the ratio of the number of shortest paths between any two nodes s, t in G that pass through v , to the total number of shortest paths between any two nodes in G . Let $\Psi(v)$ be the set of ordered pairs (s, t) in $V \times V$, so that s, t and v are distinct. Then,

$$C_B(v, G) = \sum_{(s,t) \in \Psi(v,G)} \frac{\sigma(s, t|v, G)}{\sigma(s, t|G)},$$

where $\sigma(s, t|G)$ is the total number of s, t -shortest paths in G , and $\sigma(s, t|v, G)$ is the number of shortest paths from s to t in G that pass through node v . The *average node betweenness* $C_B(S_C)$ of a subgraph S_C is the average node betweenness of all nodes $v \in V_C$ in the subgraph S_C ,

$$C_B(S_C) = \frac{1}{|V_C|} \sum_{v \in V_C} C_B(v, S_C).$$

3. **Average node closeness:** The normalized *closeness* of a node $v \in V$ is defined as

$$C_{Clo}(v, G) = \frac{1}{|V| - 1} \left(\sum_{w \in V, w \neq v} d(v, w|G) \right)^{-1},$$

where $d(v, w|G)$ is the length of the shortest path between two nodes $v, w \in V$. The *average node closeness* $C_{Clo}(S_C)$ of a subgraph $S_C = (V_C, E_C)$ is

$$C_{Clo}(S_C) = \frac{1}{|V_C|} \sum_{v \in V_C} C_{Clo}(v, S_C).$$

4. **Average eigenvector centrality:** Let $A = (a_{ij})$ be the adjacency matrix of $G = (V, E)$, i.e., A is a symmetric $|V| \times |V|$ matrix with entry $a_{ij} = 1$ if $v_i, v_j \in E$ and $a_{ij} = 0$ otherwise. The *eigenvector centrality* C_E of a node $v \in V$ is

$$C_E(v, G) = \frac{1}{\lambda} \sum_{w \in V} a_{w,v} C_E(w, G),$$

where λ is the (absolute) largest eigenvalue of A . The *average eigenvector centrality* $C_E(S_C)$ for a subgraph $S_C = (V_C, E_C)$ is defined as

$$C_E(S_C) = \sum_{v \in V_C} \frac{1}{|V_C|} C_E(v, S_C).$$

Eigenvector centrality is based on the idea that importance of a node is determined by the importance of its neighbors: a node becomes more important the more important its neighbors are.

5. **Average clustering coefficient:** Let $N_v = \{w \in V : (v, w) \in E\}$ be the set of all neighbors of a node $v \in V$. The local clustering coefficient of v is then defined as

$$C_{Clu}(v, G) = \frac{|\{e_{j,k} \in E : j, k \in N_v\}|}{|N_v|(|N_v| - 1)/2}.$$

For a given subgraph $S_C = (V_C, E_C)$, we define the *average clustering coefficient* $C_{Clu}(S_C)$ as the mean of $C_{Clu}(v, S_C)$ over all $v \in V_C$.

6. **Mean path length:** The *mean path length* for a subgraph $S_C = (V_C, E_C)$ is the average length of all shortest paths between all pairs of nodes $s, t \in V_C$ in the graph S_C :

$$C_P(S_C) = \frac{1}{|V_C|(|V_C| - 1)} \sum_{s,t \in V_C} d(s, t|S_C),$$

where $d(s, t|S_C)$ is the length of the shortest path between nodes s and t in the subgraph S_C .

In addition to the network centrality measures above, we also used the following similarity coefficients to filter clusters:

1. **Dice similarity coefficient:** For any given node $v \in V$ in a graph G , let $E_v^G := \{e_{v,w} \in E\}$ be the set of edges adjacent to v . The dice similarity coefficient of the edge sets E_v^G and E_w^G of two nodes $v, w \in V$ is defined as

$$C_{DS}(v, w, G) = \frac{2|E_v^G \cap E_w^G|}{|E_v^G| + |E_w^G|}.$$

The average dice similarity of a subnetwork $S_C = (V_C, E_C)$, $V_C \subseteq V$, is

$$C_{DS}(S_C) = \frac{2}{|V_C|(|V_C| - 1)} \sum_{v,w \in V_C} C_{DS}(v, w, S_C).$$

2. **Wang similarity coefficient:** This coefficient is biologically motivated and is based on similarity between gene ontology terms. Wang similarity takes the hierarchical structure of the GO graph into account by aggregating the information of ancestor terms when comparing two GO annotations [56]. Writing $C_G(v, w)$ for the Wang similarity between the GO annotations of nodes v and w , we compute the within-cluster similarity $C_G(S_C)$ as the average Wang similarity $C_G(v, w)$ between all pairs of genes v, w in the subnetwork S_C .

We note that a number of different measures have been proposed to compute the semantic similarity between two GO terms, for a comprehensive review see Pesquita

et al. [57]. The choice of GO semantic similarity measure and a comparative evaluation of different measures are still subject to debate in the literature, as no gold standard exists, and different studies come to different conclusions [57]. The choice of similarity measure is therefore somewhat arbitrary and a matter of personal preferences. We opted for Wang similarity because of own good experiences with this coefficient in previous work, and because it is implemented in the *GOSemSim* package in R [58], which helped seamless integration into our analysis script. We note however that Wang similarity can easily be replaced by other semantic similarity measures in our analysis pipeline.

Filtering of clusters in C^{hit} was performed using the above topological and similarity measures as follows: We computed all topological and similarity measures for each subnetwork in C^{all} , and performed a Wilcoxon test to assess differences of means of significantly enriched subnetworks in C^{hit} with randomly selected clusters in $C^{all} \setminus C^{hit}$ of the same size. Clusters that yielded a significant difference of the mean for all or all but one topological and semantic similarity measure at a significance level of 5% were considered for further analysis. By this, we ensure a stringent selection of subnetworks for further analysis: Resulting subnetworks are both enriched with hits from the RNAi screens, and show topological properties that distinguish them from random clusters. In combination, these criteria resulted in a stringent selection of subnetworks, compare Table 1. We note that in theory, due to the variation of the cluster size parameter in ClusterONE, C^{hit} may contain clusters that are subsets/supersets of one another, however after filtering using the similarity and centrality measures we did not observe clusters that

were subsets or supersets of other clusters in the analysis performed here.

Software and availability:

We implemented our data analysis pipeline in R [59]. Graph based calculations and reconstruction of subnetworks were performed using the *iGraph* library [60]. Network visualization was performed using *Cytoscape* [61]. All Reactome pathway and GO based enrichments were computed using the Bioconductor packages *clusterProfiler* and *ReactomePA* [62,63]. Semantic similarities were computed using the *GOSemSim* package [58]. R-code and data used are available on request from the authors.

Results

Given the long and often largely non-overlapping hit lists from RNAi screens targeting viral infection, a central aim of our analysis was to select a small number of most significant, infection-relevant host protein subnetworks for further manual analysis, and thus to pick most promising candidates from the original screens for functional characterization. We are therefore interested in a small set of significant clusters, that are both enriched with hits from the RNAi screens, and play a central role in the host or virus-host protein interaction network.

We used RNAi data from seven different, published genome-wide RNAi screens focusing on the three viruses HIV [10,12,53], HCV [13,18,54] and WNV [11]. Hit lists from screens targeting the same virus were combined and analyzed in a virus-specific way, as well as all data pooled for pan-viral analysis of host restriction and host dependency factors. Data were analyzed as described in

Table 1 P-values of Wilcoxon test to determine significance of mean values of network centralities and semantic measures for subnetwork

Centrality measure	HIV		HCV		Combined		
	s66	s52	s43	s64	s46	s52	s239
Betweenness	< 0.0001	0.0131	0.0247	0.0005	0.0131	0.0131	0.0040
Closeness	< 0.0001	0.0131	0.0247	0.0005	0.0131	0.0131	0.0040
Clustering Coefficient	< 0.0001	0.0247	0.0001	0.0005	0.0131	0.0131	0.0040
Eigenvector Centrality	< 0.0001	0.0131	1	1	1	0.0057	0.0057
Node Degree	< 0.0001	1	0.0247	0.0005	0.0211	0.0131	0.0040
Path Length	< 0.0001	0.0131	0.0247	0.0002	0.0131	0.0131	0.0040
Dice Similarity	< 0.0001	0.0131	0.0247	0.0005	0.0131	0.0131	0.0040
Wang Sim. (GO.BP)	0.0004	0.0286	0.5926	0.6009	0.0284	0.0286	0.0136
Wang Sim. (GO.CC)	0.0004	0.0286	0.5926	0.0315	0.0284	0.0286	0.0136
Wang Sim. (GO.MF)	0.0004	0.0286	0.0498	0.7713	1	0.3429	0.1077

A Wilcoxon test was used to determine the significance of network centrality measures and semantic similarity measures of subnetworks significantly enriched with RNAi screening hits. Average similarity measures over all nodes in a given enriched cluster were tested against non-enriched subnetworks of comparable size, using a Wilcoxon test to assess significance of the differences between the means for each of the given network centrality and semantic similarity measures. Shown are resulting p-values for two clusters for HIV, two clusters for HCV, and three combined clusters.

Materials and methods and as illustrated in Figure 1. Analysis of the single West Nile virus screen did not yield significant results after filtering, probably due to too small number of hits included in the analysis. We did include this virus in the pan-viral analysis. Table 2 gives an overview over resulting hits for HIV-1 and HCV, discussed in more detail below.

Human immunodeficiency virus-1 (HIV-1)

Two significant subnetworks of size 52 (HIV_s52) and 66 proteins (HIV_s66), respectively, were obtained from analysis of the three HIV screens after filtering as described in Materials and methods. These subnetworks are shown in Additional file 1: Figure S1 and Additional file 2: Figure S2, respectively. A Reactome pathway enrichment analysis of the subnetworks as well as the original screens is shown in Figure 2A. The pathway analysis of the three screens individually yields the expected, albeit very general pathways, such as *Immune System*, *HIV Infection*, *Metabolism* or *Signal Transduction*. This is a typical outcome for geneset or pathway enrichment analysis with large hit lists from RNAi screens, which often results in very unspecific and general terms as the only significant

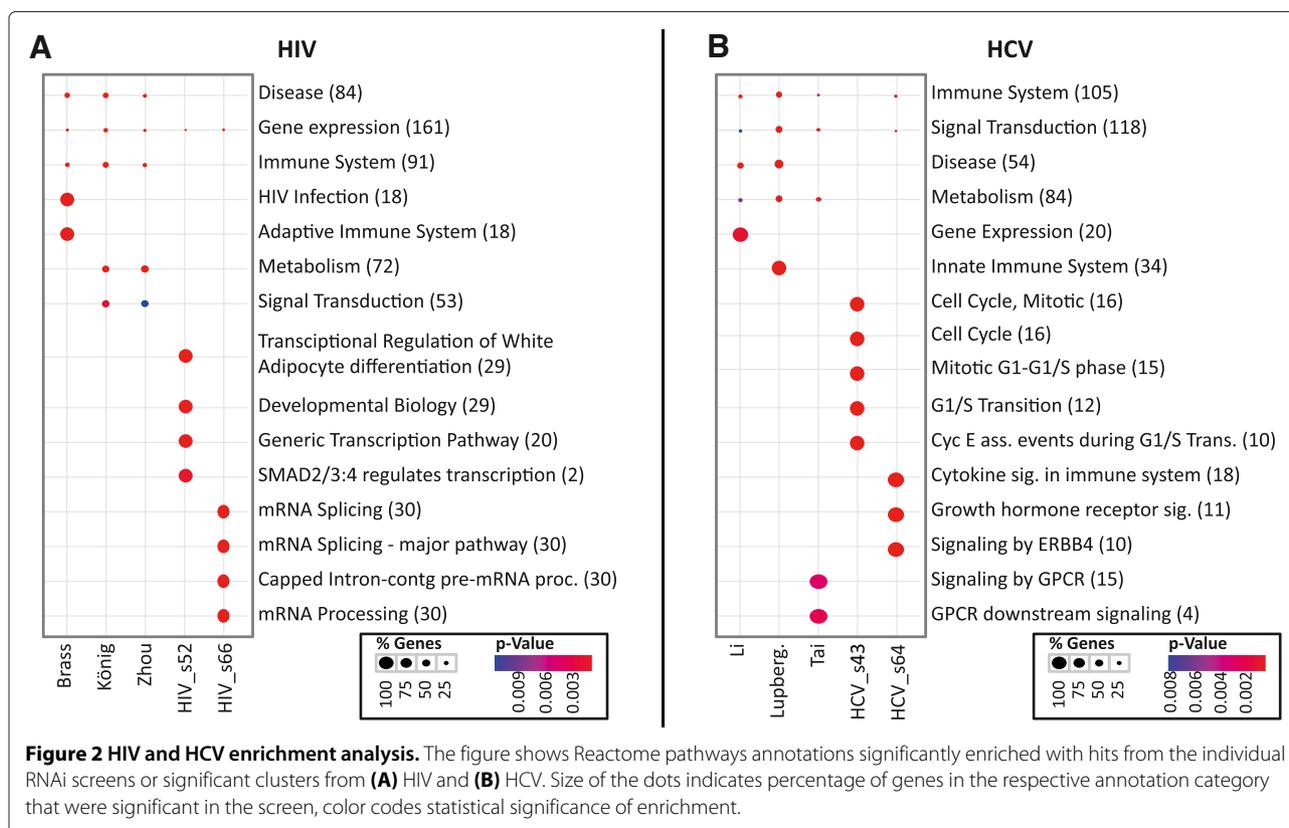
outcomes. In contrast, due to the inclusion of protein neighborhoods and focusing on enriched subnetworks of the host protein network, much more specific results can be obtained using our approach, as illustrated for the HIV_s52 and HIV_s66 subnetworks (Figure 2A).

The HIV_s52 subnetwork consists primarily of genes involved in transcription, and comprises in particular subunits of the mediator complex. This complex is a transcriptional coactivator, involved in the regulation of expression of RNA polymerase II transcripts, and thus of all protein coding and most non-coding RNA genes [64]. The mediator complex has previously been identified in the context of HIV-1 infection in the meta-analysis by Bushman et al. [24] and was a major hit in the RNAi screens by Zhou et al. [53] and König et al. [12]. This discovery has led to different hypotheses about the role of the mediator complex in HIV infection. While Zhou et al. suggest that mediator complex subunits are required for Tat-activated transcription, König et al. speculate that the complex may be involved in reverse transcription. The exact role of the mediator complex in the HIV lifecycle still needs to be determined. Interestingly, transcriptional regulation does not show up in individual enrichment

Table 2 Key results achieved for HIV-1 and HCV

Virus	Subnetwork	Predicted novel host factors
HIV	HIV_s52	<ul style="list-style-type: none"> • KDM4B - lysine-specific demethylase 4B
	HIV_s66	<ul style="list-style-type: none"> • HNRNPK - Heterogeneous nuclear ribonucleoprotein K (hnRNP K) (Transformation up-regulated nuclear protein) (TUNP) • HNRNPL - Heterogeneous nuclear ribonucleoprotein L • HNRNPM - Heterogeneous nuclear ribonucleoprotein M • HNRNPU - Heterogeneous nuclear ribonucleoproteinU (hnRNP U) (Scaffold attachment factor A) (SAF-A) (p120) (pp120) • RBM11 - Splicing regulator RBM11 (RNA-binding motif protein 11) • RBM41 - RNA-binding protein 41 (RNA-binding motif protein 41) • RBM42 - RNA-binding protein 42 (RNA-binding motif protein 42) • RBM4B - RNA-binding protein 4B (RNA-binding motif protein 30) (RNA-binding motif protein 4B) (RNA-binding protein 30) • RBM7 - RNA-binding protein 7 (RNA-binding motif protein 7) • SRSF3 - Serine/arginine-rich splicing factor 3 (PremRNA-splicing factor SRP20) (Splicing factor, arginine/serine-rich 3), • SRSF4 - Serine/arginine-rich splicing factor 4 (Pre-mRNA-splicing factor SRP75) (SRP001LB) (Splicing factor, arginine/serine-rich 4) • SRSF10 - Serine/arginine-rich splicing factor 10 (40 kDa SR-repressor protein)
HCV	HCV_s43	<ul style="list-style-type: none"> • $\alpha\beta$ Crystallin Complex subunits (CRYBAA, CRYBAB, CRYBA1, CRYBA2, CRYBA4, CRYBA1, CRYBB1, CRYBB2, CRYBB3) • Heat-shock proteins (HspB1, HspB2, HspB6, HspB7 and HspB8)
	HCV_s64	<ul style="list-style-type: none"> • Tyrosine-protein phosphatase non-receptors, various types (PTP-1B, TCPTP, PTP-H1, PTPase MEG2) • Tankyrase-1 (Poly-ADP-ribosyltransferase)

The table shows the main novel findings for HIV-1 and hepatitis C virus obtained by mapping RNAi data to protein interaction networks, and using the clustering and filtering procedure proposed here. Results for the combined analysis are given in Additional file 5.



analysis of the screens by König et al. and Zhou et al. In contrast, it is highly significant for the HIV_s52 sub-network, underlining the gain in power brought by a meta-analysis and by inclusion of protein neighborhoods in analyzing RNAi data (Figure 2).

The HIV_s66 subnetwork comprises many members of the heterogeneous nuclear ribonucleoprotein subunits (hnRNP) and serine/arginine rich splicing factors. The different hnRNP subunits participate in different steps in the RNA metabolism, including splicing, export, localization and translation [65]. Similarly, several of the serine/arginine rich splicing factors in the HIV_s66 sub-network are known to have direct interactions with HIV viral proteins [66]. Correspondingly, enriched pathways in the HIV_s66 subnetwork are related to mRNA processing and splicing (Figure 2A). A recent study by Lund et al. focused on the hnRNP complexes, and mechanistic details of its involvement in HIV-1 infection [67]. The authors report that loss of the hnRNP A1 subunit increases the expression of HIV Gag and Env, but with no subsequent increase of viral RNA. In contrast, depletion of hnRNP A2 increases both Gag protein and HIV-1 RNA levels. Changes in expression of different isoforms of hnRNP D had very diverse effects, where some isoforms increased HIV-1 gene expression, whereas others brought the cells into a non-permissive state.

Hepatitis C virus

We next repeated the analysis for the three hepatitis C virus screens by Li et al., Tai et al. and Lupberger et al. [13,18,54]. Combined analysis and submodule filtering as above resulted in two different subnetworks with 43 proteins (HCV_s43) and 64 proteins (HCV_s64), respectively, compare Additional file 3: Figure S3 and Additional file 4: Figure S4. Reactome enrichment showed that both modules were functionally very specific (Figure 2B).

The HCV_s43 module mainly contains dual specificity protein phosphatases, heat shock proteins (HSPs), crystalline proteins and mitogen-activated protein kinases (MAPKs). In particular the MAPKs are interesting, as they play a key role in cell growth and proliferation and are associated with hepatocellular carcinoma - the end stage of chronic HCV infection [68]. On the other hand, the HSPs and crystalline proteins both act as chaperones. Hsp72, one of the heat shock proteins in the HCV_s43 network, is known to be a positive regulator of HCV RNA replication by increasing replication complex levels [69]; furthermore, Lim et al. recently showed that the viral protein NS5A increases Hsp72 levels through the transcription factors HSF1 and NFAT5 [70], thus increasing its own replication. Reactome enrichment analysis of the HCV_s64 subnetwork shows enrichment in cytokine signaling, growth hormone receptor signaling, and ERBB4

signaling. The subnetwork in particular comprises several interleukin receptors and subunits, as well as insulin receptor and receptor substrate. The interleukins play an important role in suppression of infection, it is thus no surprise that HCV itself interacts with different interleukins to inhibit the cellular antiviral response [71-73].

Pan-viral host factors

To get an overview over pan-viral host factors, we next pooled all seven screens (3 HIV, 3 HCV, 1 WNV) and analyzed the combined hit list [10-13,18,53,54]. Using our pipeline, we identified three highly significant subnetworks of size 46 proteins (Combi_s46), 52 proteins (Combi_s52) and a large network with 239 proteins (Combi_s239). The Combi_s52 network was identical to the one described for HIV, and is thus not discussed further here (see results on HIV).

The Combi_s239 subnetwork contains 17 tyrosine-protein kinases, 6 tyrosine-protein phosphatase non-receptors, 5 insulin receptor substrates, and an insulin receptor (see Additional file 5 and Figure 3). Indeed, insulin resistance is one of the effects observed in HCV infected patients as the disease progresses. A recent study identified components of the insulin signaling pathway that are altered by HCV, conferring insulin resistance in the patient [74]. The study showed that PTPB1, a tyrosine phosphatase, is significantly induced in infected cells. Supporting evidence also comes from a study by Garcia-Ruiz et al. who showed that insulin resistance is also associated with IFN- α resistance in Hep-G2 cells with increase PTPB activity [75]. Both these resistance types were lowered using Metformin, in both studies. The presence of several PTPBs in this network provides a basis for further experimentation with appropriate drugs that can keep the insulin-IFN- α resistances in check.

The Combi_s239 subnetwork furthermore contains several proteins from the Src kinase family. In WNV, it is known that e.g. c-Yes, a member of this family, is required for transportation of virions through the secretory pathway [76]. Several of the Src kinase family members are activated by HIV Nef [77], and also HCV NS5A induces phosphorylation events in the Src family [78-80].

The Combi_s46 subnetwork consists primarily of SMAD and zinc finger proteins. The SMADs are involved in TGF- β signaling, where they activate downstream gene expression [81,82]. TGF- β is an immunosuppressive cytokine, its modulation is therefore advantageous for parasitic viruses [83,84]. Indeed, HCV suppresses the TGF- β mediated transcriptional activation by the full-length polyprotein and NS3-viral proteins in a SMAD-R dependent manner [85]. Zinc finger proteins on the other hand have antiviral activity: Sakkhachornphop et al. have shown that a zinc-finger protein targets the 2-long terminal repeat (2-TLR) circle junctions of HIV-1 DNA

[86,87]. This region of the HIV genome is cleaved by HIV integrase, and blocking this site restricts HIV-1 gene transcription.

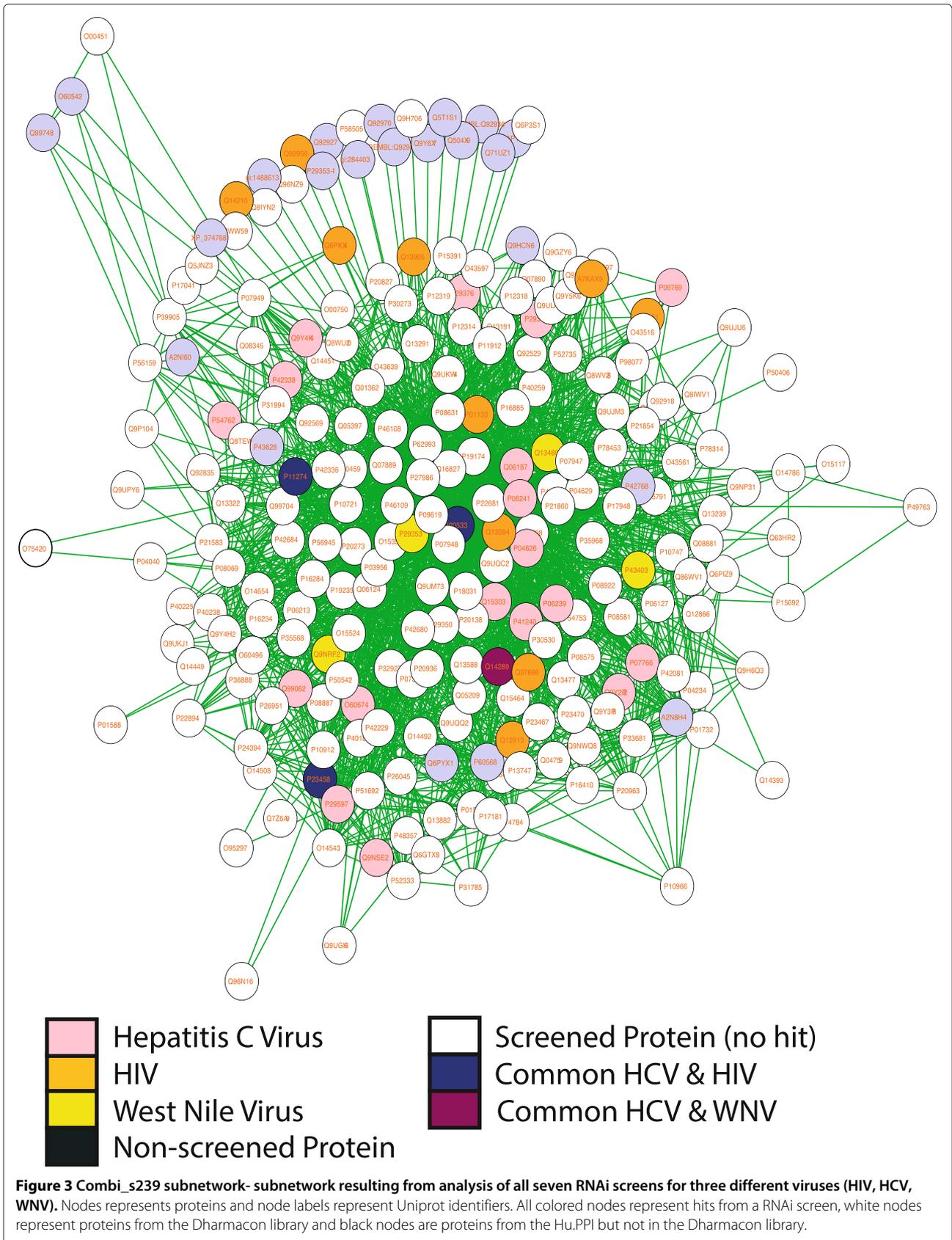
Mapping tissue-specific expression data

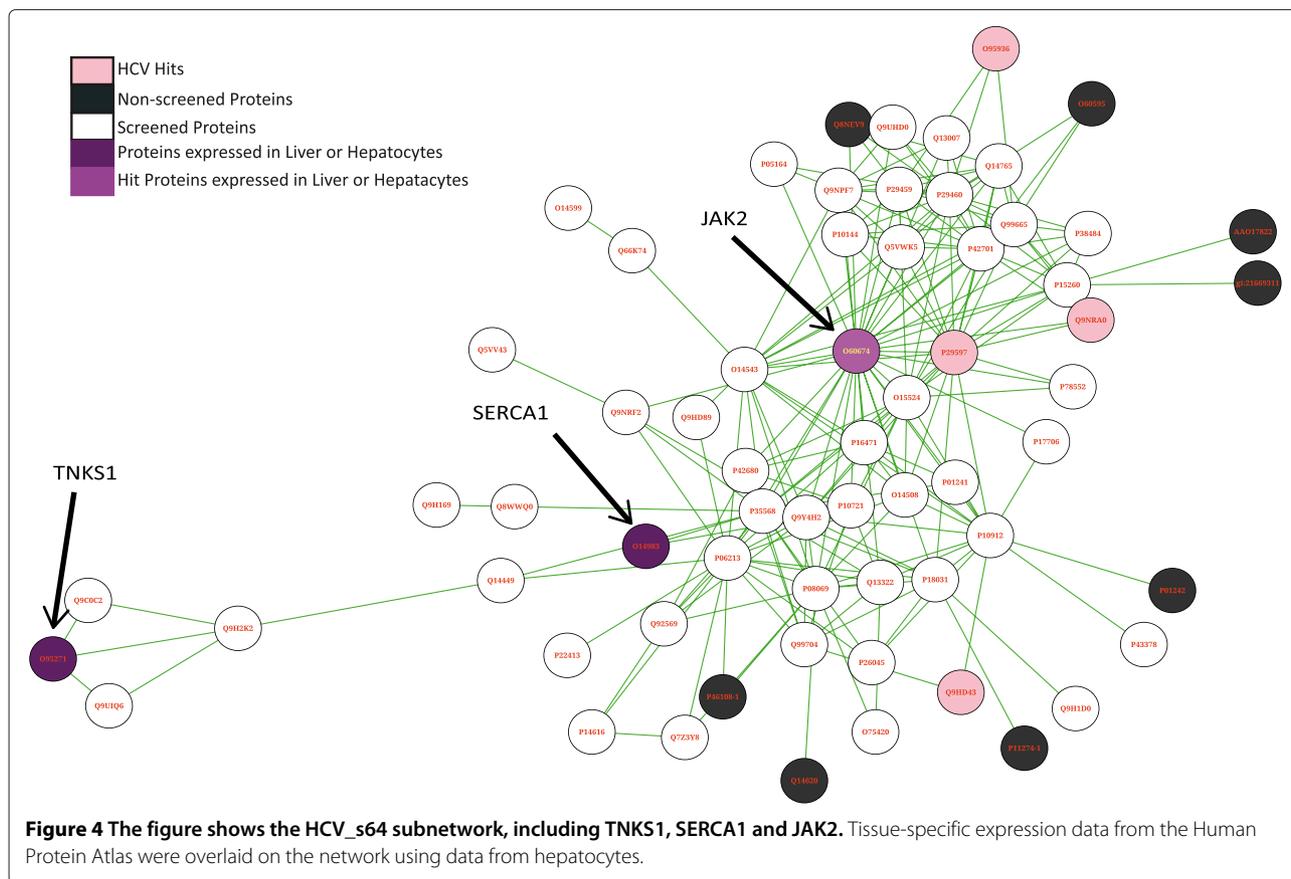
Given the filtered, significant subnetworks for the different viruses, we next addressed the problem to select suitable candidates for further experimental validation from the subnetworks, and thus ultimately possible targets for antiviral drugs. Of particular interest are proteins that are strongly expressed in tissues targeted by a given virus. Such tissue-specific or cell-line specific expression data is widely available through the Human Protein Atlas [88]. We overlaid subnetworks with tissue-specific expression data, and retained only proteins in the subnetwork that had moderate or high expression levels in the Protein Atlas database. Given the high rates of false negatives in RNAi screens [27], we do not necessarily require that candidate genes are direct hits in any of the screens.

For hepatitis C virus, expression levels were selected from hepatocytes, resulting in three proteins that remained in the HCV-s64 subnetwork: Tankyrase-1 (TNKS1, also known as PARP5A, PARPL, TIN1 and TINF1), Sarcoplasmic/endoplasmic reticulum calcium ATPase 1 (SERCA1) and JAK2, compare Figure 4. Of these, TNKS1 and SERCA1 have not been reported as hits in any of the three HCV screens used. Interestingly, SERCA2, a close family member of SERCA1, has been shown to play an important role in HCV core induced ER stress and control of apoptosis [89]. As SERCA1 is closely interacting with SERCA2 and has similar functions, a similar role might be played by SERCA1 in HCV infection. TNKS1 on the other hand is involved in WNT signaling, regulation of telomere length, and vesicle trafficking. TNKS1 has previously been suggested as an attractive anti-cancer target [90], and is involved in HCV-induced apoptosis [91]. In case of HIV, we filtered proteins based on expression in macrophages. This resulted mainly in different subunits of the heterogeneous nuclear ribonucleoproteins (hnRNPs) as highly expressed putative antiviral targets.

Discussion and conclusion

Genome wide RNAi screening experiments typically result in lists of hundreds of "hit" genes, and the selection of promising candidates for biochemical follow-up as well as their placement in the underlying molecular processes is a significant challenge [20]. To complicate matters further, in particular for viral RNAi screens, very low overlap has been reported even for screens targeting the same virus [24]. High false negative rates are likely a major contributing factor to this problem [27]. While geneset enrichment approaches can help to interpret lists of hit genes, they in our experience typically lead to very





general, unspecific terms and often fail to achieve statistical significance for concrete, specific biological processes or pathways when applied to RNAi screening data. This problem clearly is aggravated if hit lists are prone to high levels of false negative results, and it is then a very challenging problem to pick interesting candidates for further experimental characterization.

In this work, we have developed a network-based approach for gene prioritization. The simple underlying idea is to interpret hit genes from RNAi screening experiments in their biological context, by taking the host cell protein-protein interaction (PPI) network into account. We cluster this PPI network to identify highly connected subnetworks, and then map the RNAi data onto this clustered network to find enriched submodules. Additional experimental data such as known virus-host interactions, gene expression data or e.g. proteomics data can easily be integrated at this stage and can be included in the network-based analysis. Similarly, it is straightforward to combine data from different screens for the same or even for different viruses at this level, to enable a network-based meta analysis of virus-host interactions. We exemplify this in a meta-analysis over seven different viral RNAi screens targeting three different viruses. In contrast to traditional geneset enrichment analysis, no

prior definition of relevant gene sets (e.g. gene ontology annotations or biological pathways) is required, but instead gene sets are automatically defined by clustering of the PPI network. This is indeed an advantage and disadvantage at the same time: While we do not require a-priori defined gene sets for our analysis, our approach clearly depends on the underlying PPI network that must be given as input. Unfortunately, in particular for yeast-2-hybrid experiments, such networks are known to contain many false positive connections, which may negatively impact our analysis. Furthermore, we specifically opted to include high-confidence predicted interactions from the STRING database, which was required to obtain a sufficiently dense, connected network to permit further analysis. There is thus an inherent tradeoff between reliability of the underlying network used and sufficient network size and connectivity to allow a meaningful analysis. Similarly, the choice of clustering algorithm and similarity measures used to further filter significant networks will impact results. As proteins often perform multiple functions in a cell, we decided to use a clustering algorithm that allows for overlaps between different clusters, permitting individual proteins to be part of several different subnetworks. We furthermore performed our analysis with a whole range of parameters for the desired cluster

size, using a redundant set of clusters of different sizes in the ensuing network centrality and similarity based filtering step. We thereby let the algorithm automatically select significant clusters of all sizes.

As no gold standard is available for virus-host interaction networks and RNAi screening data analysis, it is very difficult to assess the influence these different clustering parameters and false-positive or false-negative interactions in the underlying PPI network have on results. Reassuringly, our results show that we recover many of the known hits for the different viruses used in this study, and top candidates resulting from our gene prioritization approach are largely confirmed by other meta-analysis approaches that have been performed using different methods. For example, Bushman et al. performed a meta-analysis of all published HIV-1 RNAi screens in 2009 [24], and also identified the mediator complex and hnRNPs as major HIV-1 host cell factors in their analysis. The mediator complex is also reported by Murali et al. in their analysis [38], whereas two further studies by Bader and Nepusz, respectively, identified the hnRNPs using MCODE, a different clustering algorithm than employed in our work [55,92]. Other related approaches include the work by MacPherson et al. [39], Dickerson et al. [30], Snijder et al. [37] and the VirHostNet database developed by Navratil et al. [93]. A unique aspect of our analysis is the comparative analysis over different viruses, with a specific focus on functional subnetworks in this pan-viral meta-analysis.

There are two further assumptions that we make in our analysis, that are worthy a brief discussion. The first, noncritical assumption we made in this manuscript concerns the expression analysis, overlaying the tissue specific expression data for hit selection onto the PPI network. We here made the assumption that low tissue expression of a gene implies that the gene is not a good target and was used as reason to exclude the gene from further consideration. We use this assumption here to filter genes within a subnetwork, but this is clearly a very crude approximation and many cases are conceivable where also a lowly expressed gene may be a very good drug target and may play an important role in infection. Obviously the inverse is not true: High expression alone does not make a gene a good target. The second assumption is critical: Our subnetwork analysis is based on the assumption that due to technical and biological variability, different genes within a subnetwork may be identified in different screens, but that indeed the entire subnetwork or sub-complex is a relevant host factor. In particular in light of high false negative rates in RNAi screens [27] and further variability due to e.g. different experimental protocols, cell lines and viral genotypes used and different transfection and infection times, it is very plausible that different genes in the same pathway or subnetwork will be identified in different

screens, even when targeting the same virus. Our further subnetwork analysis therefore requires that subnetworks resulting from the clustering have high functional consistency, in the sense that the proteins within one cluster need to be involved in the same biological process or pathway, whereas different clusters should be functionally distinct – this is a *conditio sine qua non* when speaking of significance of a subnetwork. In line with this, the identification of putative targets in our analysis focuses on all proteins in a subnetwork, even if they did not show up as hits in any of the original screens considered. Before proceeding with such hits in a drug development pipeline, clearly additional experiments are required to confirm a role of these hits in the infection process, and in particular an effect of targeting the candidate gene on viral infection. As cells have many redundant mechanisms, even if a host gene is involved in viral infection, targeting this gene may not be sufficient to inhibit viral replication. Detailed mathematical modeling of the underlying processes in the subnetwork may then be a good option to identify optimal treatment strategies, but goes beyond the scope of the present work [94].

While we have developed the approach presented in this manuscript for the analysis of viral RNAi screening data, the general pipeline is applicable to any type of experiment resulting in long “hit” gene lists. Examples include gene expression data e.g. from microarray or transcriptome sequencing experiments, methylation profiles, genomic data such as array CGH or DNA sequencing, and proteomic assays based on mass spectrometry or protein arrays. Similarly, biological questions addressable with our pipeline extend well beyond viral infection, and basically include any assay where a mechanistic biological understanding is sought for based on large-scale, high-throughput data sets. In particular with the current developments in and increasing availability of big data in biology, network-based analysis approaches are a fundamental tool to interpret and understand the underlying biological processes, and will become more and more important as available data grows. We demonstrate the use of such network-based analysis methods on the concrete example of virus-host interactions in the present work.

Additional files

Additional file 1: Figure S1. HIV_s52 subnetwork: The figure shows the HIV_s52 subnetwork resulting from the analysis of the HIV screens. The subnetwork primarily consists of genes involved in transcription, and particularly comprises the mediator complex.

Additional file 2: Figure S2. HIV_s66 subnetwork: Shown is the HIV_s66 subnetwork resulting from the HIV screen analysis. The network essentially contains splicing factors and members of the hnRNP complex.

Additional file 3: Figure S3. HCV_s43 subnetwork: This subnetwork from the analysis of the three HCV screens comprises mainly heat shock proteins and proteins of the MAPK pathway.

Additional file 4: Figure S4. HCV_s64 subnetwork: The HCV_s64 subnetwork is one of two significant subnetworks for the HCV screens, and contains interleukin receptors, cytokines and growth hormone receptors.

Additional file 5: List of proteins in Combined_s239 subnetwork. This xls file contains the proteins involved in the Combined_s239 network, together with additional annotation.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

SA designed and implemented the method, analyzed data and wrote the first draft of the manuscript. LK conceived and designed the work, and wrote the final version of the paper. Both authors read and approved the final manuscript.

Acknowledgments

The authors acknowledge funding from the BMBF (GerontoSys/Agenet, grant 031A080) and the European Union (FP 7, grant 260429, SysPatho). SA was partially funded by the HGS MathComp Graduate School of Heidelberg University. We would like to thank G. Suryavanshi and N. Kiani as well as two anonymous referees for useful comments and suggestions.

Author details

¹Institute of Medical Informatics and Biometry, Medical Faculty, TU Dresden, Fetscherstr. 74, 01307 Dresden, Germany. ²ViroQuant Research Group Modeling, BioQuant, Heidelberg University, INF 267, 69120 Heidelberg, Germany. ³Present address: Department of Translational Genomics/Center for Molecular Medicine, University of Cologne, Robert-Koch Str. 21, 50931 Cologne, Germany.

Received: 24 September 2014 Accepted: 27 January 2015

Published online: 13 February 2015

References

- Boutros M, Kiger AA, Armknecht S, Kerr K, Hild M, Koch B, et al. Genome-wide RNAi analysis of growth and viability in drosophila cells. *Science*. 2004;303(5659):832–5.
- Furlong EE. A functional genomics approach to identify new regulators of Wnt signaling. *Dev Cell*. 2005;8(5):624–6. doi:10.1016/j.devcel.2005.04.006.
- Muller P, Kuttenukeuler D, Gesellchen V, Zeidler MP, Boutros M. Identification of JAK/STAT signalling components by genome-wide RNA interference. *Nature*. 2005;436(7052):871–5. doi:10.1038/nature03869.
- Friedman A, Perrimon N. A functional RNAi screen for regulators of receptor tyrosine kinase and ERK signalling. *Nature*. 2006;444(7116):230–4. doi:10.1038/nature05280.
- Kittler R, Pelletier L, Heninger AK, Slabicki M, Theis M, Miroslaw L, et al. Genome-scale RNAi profiling of cell division in human tissue culture cells. *Nat. Cell Biol*. 2007;9:1401–12.
- Chia N-Y, Chan Y-S, Feng B, Lu X, Orlov YL, Moreau D, et al. A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. *Nature*. 2010;468(7321):316–20. doi:10.1038/nature09531.
- Collinet C, Stöter M, Bradshaw CR, Samusik N, Rink JC, Kenski D, et al. Systems survey of endocytosis by multiparametric image analysis. *Nature*. 2010;464(7286):243–9. doi:10.1038/nature08779.
- Ebert AD, Laussmann M, Wegehingel S, Kaderali L, Erfle H, Reichert J, et al. Tec-kinase-mediated phosphorylation of fibroblast growth factor 2 is essential for unconventional secretion. *Traffic*. 2010;11(6):813–26. doi:10.1111/j.1600-0854.2010.01059.x.
- Theis M, Buchholz F. High-throughput RNAi screening in mammalian cells with esirnas. *Methods*. 2011;53(4):424–9. doi:10.1016/j.ymeth.2010.12.021.
- Brass AL, Dykxhoorn DM, Benita Y, Yan N, Engelman A, Xavier RJ, et al. Identification of host proteins required for HIV infection through a functional genomic screen. *Science*. 2008;319(5865):921–6. doi:10.1126/science.1152725.
- Krishnan MN, Ng A, Sukumaran B, Gilfoy FD, Uchil PD, Sultana H, et al. RNA interference screen for human genes associated with west Nile virus infection. *Nature*. 2008;455(7210):242–5. doi:10.1038/nature07207.
- König R, Zhou Y, Elleder D, Diamond TL, Bonamy GMC, Ireland JT, et al. Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication. *Cell*. 2008;135(1):49–60. doi:10.1016/j.cell.2008.07.032.
- Tai AW, Benita Y, Peng LF, Kim S-S, Sakamoto N, Xavier RJ, et al. A functional genomic screen identifies cellular cofactors of hepatitis c virus replication. *Cell Host Microbe*. 2009;5(3):298–307. doi:10.1016/j.chom.2009.02.001.
- Börner K, Hermle J, Sommer C, Brown NP, Knapp B, Glass B. From experimental setup to bioinformatics: an RNAi screening platform to identify host factors involved in hiv-1 replication. *Biotechnol J*. 2010;5(1):39–49. doi:10.1002/biot.200900226.
- Karlas A, Machuy N, Shin Y, Pleissner K-P, Artarini A, Heuer D, et al. Genome-wide RNAi screen identifies human host factors crucial for influenza virus replication. *Nature*. 2010;463(7282):818–22. doi:10.1038/nature08760.
- König R, Stertz S, Zhou Y, Inoue A, Hoffmann H-H, Bhattacharyya S, et al. Human host factors required for influenza virus replication. *Nature*. 2010;463(7282):813–7. doi:10.1038/nature08699.
- Reiss S, Rebhan I, Backes P, Romero-Brey I, Erfle H, Matula P, et al. Recruitment and activation of a lipid kinase by hepatitis c virus NSSA is essential for integrity of the membranous replication compartment. *Cell Host Microbe*. 2011;9(1):32–45. doi:10.1016/j.chom.2010.12.002.
- Lupberger J, Zeisel MB, Xiao F, Thumann C, Fofana I, Zona L, et al. EGFR and EphA2 are host factors for hepatitis c virus entry and possible targets for antiviral therapy. *Nat Med*. 2011;17(5):589–95. doi:10.1038/nm.2341.
- Metz P, Dazert E, Ruggieri A, Mazur J, Kaderali L, Kaul A, et al. Identification of type i and type ii interferon-induced effectors controlling hepatitis c virus replication. *Hepatology*. 2012;56(6):2082–93. doi:10.1002/hep.25908.
- Moffat J, Sabatini DM. Building mammalian signalling pathways with RNAi screens. *Nat Rev Mol Cell Biol*. 2006;7:177–87.
- Kaderali L, Dazert E, Zeuge U, Frese M, Bartenschlager R. Reconstructing signaling pathways from RNAi data using probabilistic Boolean threshold networks. *Bioinformatics*. 2009;25:2229–35.
- Houzet L, Jeang K-T. Genome-wide screening using RNA interference to study host factors in viral replication and pathogenesis. *Exp Biol Med*. 2011;236(8):962–7. doi:10.1258/ebm.2010.010272. Accessed 2013-02-17.
- Mohr S, Bakal C, Perrimon N. Genomic screening with RNAi: results and challenges. *Annu Rev Biochem*. 2010;79:37–64. doi:10.1146/annurev-biochem-060408-092949.
- Bushman FD, Malani N, Fernandes J, D'Orso I, Cagney G, Diamond TL, et al. Host cell factors in HIV replication: meta-analysis of genome-wide studies. *PLoS Pathog*. 2009;5(5):1000437. doi:10.1371/journal.ppat.1000437.
- Snijder B, Sacher R, Ramo P, Damm EM, Liberali P, Pelkmans L. Population context determines cell-to-cell variability in endocytosis and virus infection. *Nature*. 2009;461:520–3.
- Knapp B, Rebhan I, Kumar A, Matula P, Kiani NA, Binder M, et al. Normalizing for individual cell population context in the analysis of high-content cellular screens. *BMC Bioinformatics*. 2011;12:485. doi:10.1186/1471-2105-12-485.
- Hao L, He Q, Wang Z, Craven M, Newton MA, Ahlquist P. Limited agreement of independent rna screens for virus-required host genes owes more to false-negative than false-positive factors. *PLoS Comput Biol*. 2013;9(9):1003235. doi:10.1371/journal.pcbi.1003235.
- de Chasse B, Meyniel-Schicklin L, Aublin-Gex A, André P, Lotteau V. Genetic screens for the control of influenza virus replication: from meta-analysis to drug discovery. *Mol Biosyst*. 2012;8(4):1297–303. doi:10.1039/c2mb05416g.
- Dyer MD, Murali TM, Sobral BW. The landscape of human proteins interacting with viruses and other pathogens. *PLoS Pathog*. 2008;4(2):32. doi:10.1371/journal.ppat.0040032. Accessed 2012-08-20.
- Dickerson JE, Pinney JW, Robertson DL. The biological context of HIV-1 host interactions reveals subtle insights into a system hijack. *BMC Syst Biol*. 2010;4:80. doi:10.1186/1752-0509-4-80.
- van Dijk D, Ertaylan G, Boucher CA, Sloot PM. Identifying potential survival strategies of HIV-1 through virus-host protein interaction networks. *BMC Syst Biol*. 2010;4:96. doi:10.1186/1752-0509-4-96.
- Navratil V, de Chasse B, Meyniel L, Pradezynski F, André P, Rabourdin-Combe C, et al. System-level comparison of protein-protein

- interactions between viruses and the human type I interferon system network. *J Proteome Res.* 2010;9(7):3527–36. doi:10.1021/pr100326j.
33. Gulbahce N, Yan H, Dricot A, Padi M, Byrdsong D, Franchi R, et al. Viral perturbations of host networks reflect disease etiology. *PLoS Comput Biol.* 2012;8(6):1002531. doi:10.1371/journal.pcbi.1002531.
 34. Khadka S, Vangeloff AD, Zhang C, Siddavatam P, Heaton NS, Wang L, et al. A physical interaction network of dengue virus and human proteins. *Mol Cell Proteomics.* 2011;10(12):111–012187. doi:10.1074/mcp.M111.012187.
 35. Meliopoulou VA, Andersen LE, Birrer KF, Simpson KJ, Lowenthal JW, Bean AG, et al. Host gene targets for novel influenza therapies elucidated by high-throughput RNA interference screens. *FASEB J.* 2012 Apr;26(4):1372–86. doi:10.1096/fj.11-193466.
 36. Amberkar S, Kiani N, Bartenschlager R, Alvisi G, Kaderali L. High-throughput RNA interference screens integrative analysis: Towards a comprehensive understanding of the virus-host interplay. *World J Virol.* 2013;2(2):18–31.
 37. Snijder B, Sacher R, Rämö P, Liberali P, Mench K, Wolftrum N, et al. Single-cell analysis of population context advances RNAi screening at multiple levels. *Mol Syst Biol.* 2012;8:579. doi:10.1038/msb.2012.9.
 38. Murali TM, Dyer MD, Badger D, Tyler BM, Katze MG. Network-based prediction and analysis of HIV dependency factors. *PLoS Comput Biol.* 2011;7(9):1002164. doi:10.1371/journal.pcbi.1002164.
 39. MacPherson JI, Dickerson JE, Pinney JW, Robertson DL. Patterns of HIV-1 protein interaction identify perturbed host-cellular subsystems. *PLoS Comput Biol.* 2010;6(7):1000863. doi:10.1371/journal.pcbi.1000863.
 40. Maulik U, Mukhopadhyay A, Bhattacharyya M, Kaderali L, Brors B, Bandyopadhyay S, et al. Mining Quasi-Bicliques from HIV-1–Human Protein Interaction Network: A Multiobjective Biclustering Approach 2012. doi:6073AA69-DDD1-4FED-9839-7E52934E2BB2.
 41. Razick S, Magklaras G, Donaldson IM. iRefIndex: a consolidated protein interaction database with provenance. *BMC Bioinformatics.* 2008;9:405. doi:10.1186/1471-2105-9-405.
 42. Xenarios I, Rice DW, Salwinski L, Baron MK, Marcotte EM, Eisenberg D. DIP: the database of interacting proteins. *Nucleic Acids Res.* 2000;28(1):289–91.
 43. Aranda B, Achuthan P, Alam-Faruque Y, Armean I, Bridge A, Derow C, et al. The intact molecular interaction database in 2010. *Nucleic Acids Res.* 2010;38:525–31. doi:10.1093/nar/gkp878.
 44. Chatr-aryamontri A, Ceol A, Palazzi LM, Nardelli G, Schneider MV, Castagnoli L, et al. MINT: the molecular interaction database. *Nucleic Acids Res.* 2007;35:572–4. doi:10.1093/nar/gkl950.
 45. Stark C, Breitkreutz B-J, Chatr-Aryamontri A, Boucher L, Oughtred R, Livstone MS, et al. The BioGRID interaction database: 2011 update. *Nucleic Acids Res.* 2011;39:698–704. doi:10.1093/nar/gkq1116.
 46. Bader GD, Betel D, Hogue CWV. BIND: the biomolecular interaction network database. *Nucleic Acids Res.* 2003;31(1):248–50.
 47. Ruepp A, Waegle B, Lechner M, Brauner B, Dunger-Kaltenbach I, Fobo G, (Database issue). CORUM: the comprehensive resource of mammalian protein complexes–2009. *Nucleic Acids Res.* 2010;38:497–501. doi:10.1093/nar/gkp914.
 48. Güldener U, Münsterkötter M, Oesterheld M, Pagel P, Ruepp A, Mewes H-W, et al. MPact: the MIPS protein interaction resource on yeast. *Nucleic Acids Res.* 2006;34:436–41. doi:10.1093/nar/gkj003.
 49. Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, et al. Human protein reference database–2009 update. *Nucleic Acids Res.* 2009;37:767–72. doi:10.1093/nar/gkn892.
 50. Pagel P, Kovac S, Oesterheld M, Brauner B, Dunger-Kaltenbach I, Frishman G, et al. The MIPS mammalian protein-protein interaction database. *Bioinformatics.* 2005;21(6):832–4. doi:10.1093/bioinformatics/bti115.
 51. Brown KR, Jurisica I. Online predicted human interaction database. *Bioinformatics.* 2005;21(9):2076–82. doi:10.1093/bioinformatics/bti273.
 52. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguéz P, et al. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.* 2011;39:561–68. doi:10.1093/nar/gkq973.
 53. Zhou H, Xu M, Huang Q, Gates AT, Zhang XD, Castle JC, et al. Genome-scale RNAi screen for host factors required for HIV replication. *Cell Host Microbe.* 2008;4(5):495–504. doi:10.1016/j.chom.2008.10.004.
 54. Li Q, Brass AL, Ng A, Hu Z, Xavier RJ, Liang TJ, et al. A genome-wide genetic screen for host factors required for hepatitis C virus propagation. *Proc Natl Acad Sci U S A.* 2009;106(38):16410–5. doi:10.1073/pnas.0907439106.
 55. Nepusz T, Yu H, Paccanaro A. Detecting overlapping protein complexes in protein-protein interaction networks. *Nat Methods.* 2012;9(5):471–2. doi:10.1038/nmeth.1938.
 56. Wang JZ, Du Z, Payattakool R, Yu PS, Chen C-F. A new method to measure the semantic similarity of GO terms. *Bioinformatics.* 2007;23(10):1274–81. doi:10.1093/bioinformatics/btm087.
 57. Pesquita C, Faria D, Falcão AO, Lord P, Couto FM. Semantic similarity in biomedical ontologies. *PLoS Comput Biol.* 2009;5(7):1000443. doi:10.1371/journal.pcbi.1000443.
 58. Yu G, Li F, Qin Y, Bo X, Wu Y, Wang S. Gosemsim: an R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics.* 2010;26(7):976–8. doi:10.1093/bioinformatics/btq064.
 59. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2014. R Foundation for Statistical Computing. <http://www.R-project.org/>.
 60. Csardi G, Nepusz T. The igraph software package for complex network research. *InterJournal.* 2006;Complex Systems:1695.
 61. Smoot ME, Ono K, Ruschekinski J, Wang P-L, Ideker T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics.* 2011;27(3):431–2. doi:10.1093/bioinformatics/btq675.
 62. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* 2012;16(5):284–7. doi:10.1089/omi.2011.0118.
 63. Yu G. ReactomePA: Reactome Pathway Analysis. R package version 1.8.1. <http://www.bioconductor.org/packages/release/bioc/html/ReactomePA.html>.
 64. Poss ZC, Ebmeier CC, Taatjes DJ. The mediator complex and transcription regulation. *Crit Rev Biochem Mol Biol.* 2013;48(6):575–608. doi:10.3109/10409238.2013.840259.
 65. Dreyfuss G, Matunis MJ, Piñol-Roma S, Burd CG. hnRNP proteins and the biogenesis of mRNA. *Annu Rev Biochem.* 1993;62:289–321. doi:10.1146/annurev.bi.62.070193.001445.
 66. Fu W, Sanders-Ber BE, Katz KS, Maglott DR, Pruitt KD, Ptak RG. Human immunodeficiency virus type 1, human protein interaction database at NCBI. *Nucleic Acids Res.* 2009;37:417–22. doi:10.1093/nar/gkn708.
 67. Lund N, Milev MP, Wong R, Sanmuganatham T, Woolaway K, Chabot B, et al. Differential effects of hnRNP D/auf1 isoforms on HIV-1 gene expression. *Nucleic Acids Res.* 2012;40(8):3663–75. doi:10.1093/nar/gkr1238.
 68. Huynh H, Nguyen TTT, Chow K-HP, Tan PH, Soo KC, Tran E. Over-expression of the mitogen-activated protein kinase (MAPK) kinase (MEK) in hepatocellular carcinoma: its role in tumor progression and apoptosis. *BMC Gastroenterol.* 2003;3:19. doi:10.1186/1471-230X-3-19.
 69. Chen Y-J, Chen Y-H, Chow L-P, Tsai Y-H, Chen P-H, Huang C-YF, et al. Heat shock protein 72 is associated with the hepatitis C virus replicase complex and enhances viral RNA replication. *J Biol Chem.* 2010;285(36):28183–90. doi:10.1074/jbc.M110.118323.
 70. Lim YS, Shin KS, Oh SH, Kang SM, Won SJ, Hwang SB. Nonstructural 5a protein of hepatitis C virus regulates heat shock protein 72 for its own propagation. *J Viral Hepat.* 2012;19(5):353–63. doi:10.1111/j.1365-2893.2011.01556.x.
 71. Polyak SJ, Khabar KS, Paschal DM, Ezelle HJ, Duverlie G, Barber GN, et al. Hepatitis C virus nonstructural 5a protein induces interleukin-8, leading to partial inhibition of the interferon-induced antiviral response. *J Virol.* 2001;75(13):6095–106. doi:10.1128/JVI.75.13.6095-6106.2001.
 72. Brady MT, MacDonald AJ, Rowan AG, Mills KHG. Hepatitis C virus non-structural protein 4 suppresses Th1 responses by stimulating IL-10 production from monocytes. *Eur J Immunol.* 2003;33(12):3448–57. doi:10.1002/eji.200324251.
 73. Eisen-Vandervelde AL, Waggoner SN, Yao ZQ, Cale EM, Hahn CS, Hahn YS. Hepatitis C virus core selectively suppresses interleukin-12 synthesis in human macrophages by interfering with AP-1 activation. *J Biol Chem.* 2004;279(42):43479–86. doi:10.1074/jbc.M407640200.
 74. del Campo JA, García-Valdecasas M, Rojas L, Rojas A, Romero-Gómez M. The hepatitis C virus modulates insulin signaling pathway in vitro promoting insulin resistance. *PLoS One.* 2012;7(10):47904. doi:10.1371/journal.pone.0047904.

75. García-Ruiz I, Solís-Muñoz P, Gómez-Izquierdo E, Muñoz-Yagüe MT, Valverde AM, Solís-Herruzo JA. Protein-tyrosine phosphatases are involved in interferon resistance associated with insulin resistance in hepg2 cells and obese mice. *J Biol Chem*. 2012;287(23):19564–73. doi:10.1074/jbc.M112.342709.
76. Hirsch AJ, Medigeshi GR, Meyers HL, DeFilippis V, Früh K, Briese T, et al. The src family kinase c-yes is required for maturation of west nile virus particles. *J Virol*. 2005;79(18):11943–51. doi:10.1128/JVI.79.18.11943-11951.2005.
77. Tribble RP, Emert-Sedlak L, Smithgall TE. Hiv-1 nef selectively activates src family kinases hck, lyn, and c-src through direct sh3 domain interaction. *J Biol Chem*. 2006;281(37):27029–38. doi:10.1074/jbc.M601128200.
78. Nakashima K, Takeuchi K, Chihara K, Horiguchi T, Sun X, Deng L, et al. Hcv ns5a protein containing potential ligands for both src homology 2 and 3 domains enhances autophosphorylation of src family kinase fyn in b cells. *PLoS One*. 2012;7(10):46634. doi:10.1371/journal.pone.0046634.
79. Pfannkuche A, Büther K, Karthe J, Poenisch M, Bartenschlager R, Trilling M, et al. c-src is required for complex formation between the hepatitis c virus-encoded proteins ns5a and ns5b: a prerequisite for replication. *Hepatology*. 2011;53(4):1127–36. doi:10.1002/hep.24214.
80. Martin-Garcia JM, Luque I, Ruiz-Sanz J, Camara-Artigas A. The promiscuous binding of the fyn sh3 domain to a peptide from the ns5a protein. *Acta Crystallogr D Biol Crystallogr*. 2012;68(Pt 8):1030–40. doi:10.1107/S0907444912019798.
81. Derynck R, Zhang Y, Feng XH. Smads: transcriptional activators of tgfbeta responses. *Cell*. 1998;95(6):737–40.
82. Shi Y, Massagué J. Mechanisms of tgfbeta signaling from cell membrane to the nucleus. *Cell*. 2003;113(6):685–700.
83. Flavell RA, Sanjabi S, Wrzesinski SH, Licona-Limón P. The polarization of immune cells in the tumour environment by tgfbeta. *Nat Rev Immunol*. 2010;10(8):554–67. doi:10.1038/nri2808.
84. Chen W, Frank ME, Jin W, Wahl SM. Tgfbeta released by apoptotic t cells contributes to an immunosuppressive milieu. *Immunity*. 2001;14(6):715–25.
85. Cheng P-L, Chang M-H, Chao C-H, Lee Y-HW. Hepatitis c viral proteins interact with smad3 and differentially regulate tgfbeta/smad3-mediated transcriptional activation. *Oncogene*. 2004;23(47):7821–38. doi:10.1038/sj.onc.1208066.
86. Sakkhachornphop S, Jiranusornkul S, Kodchakorn K, Nangola S, Sirisanthana T, Tayapiwatana C. Designed zinc finger protein interacting with the hiv-1 integrase recognition sequence at 2-ltr-circle junctions. *Protein Sci*. 2009;18(11):2219–30. doi:10.1002/pro.233.
87. Sakkhachornphop S, Barbas CF3rd, Keawwichit R, Wongworapat K, Tayapiwatana C. Zinc finger protein designed to target 2-long terminal repeat junctions interferes with human immunodeficiency virus integration. *Hum Gene Ther*. 2012;23(9):932–42. doi:10.1089/hum.2011.124.
88. Uhlen M, Oksvold P, Fagerberg L, Lundberg E, Jonasson K, Forsberg M, et al. Towards a knowledge-based human protein atlas. *Nat Biotechnol*. 2010;28(12):1248–50. doi:10.1038/nbt1210-1248.
89. Benali-Furet NL, Chami M, Houel L, De Giorgi F, Vernejoul F, Lagorce D, et al. Hepatitis c virus core triggers apoptosis in liver cells by inducing er stress and er calcium depletion. *Oncogene*. 2005;24(31):4921–33. doi:10.1038/sj.onc.1208673.
90. Waaler J, Machon O, Tumova L, Dinh H, Korinek V, Wilson SR, et al. A novel tankyrase inhibitor decreases canonical wnt signaling in colon carcinoma cells and reduces tumor growth in conditional apc mutant mice. *Cancer Res*. 2012;72(11):2822–32. doi:10.1158/0008-5472.CAN-11-3336.
91. Alisi A, Arciello M, Petrini S, Conti B, Missale G, Balsano C. Focal adhesion kinase (fak) mediates the induction of pro-oncogenic and fibrogenic phenotypes in hepatitis c virus (hcv)-infected cells. *PLoS One*. 2012;7(8):44147. doi:10.1371/journal.pone.0044147.
92. Bader GD, Hogue CWV. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics*. 2003;4:2.
93. Navratil V, de Chasse B, Meyniel L, Delmotte S, Gautier C, André P, et al. VirHostNet: a knowledge base for the management and the analysis of proteome-wide virus-host interaction networks. *Nucleic Acids Res*. 2009;37:661–8. doi:10.1093/nar/gkn794.
94. Binder M, Sulaimanov N, Clausznitzer D, Schulze M, Hüber CM, Lenz SM, et al. Replication vesicles are load- and choke-points in the hepatitis c virus lifecycle. *PLoS Pathog*. 2013;9(8):1003561. doi:10.1371/journal.ppat.1003561.

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

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

Submit your manuscript at
www.biomedcentral.com/submit

