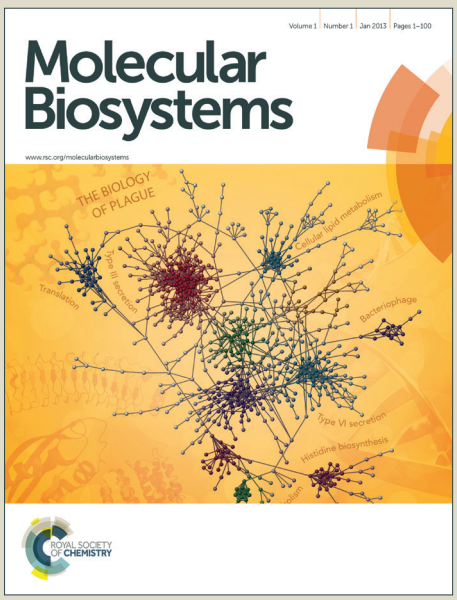
Molecular BioSystems

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/molecularbiosystems

Journal Name

RSCPublishing

ARTICLE

Viral Proteins that Bridge Unconnected Proteins and Components in Human PPI Network

Cite this: DOI: 10.1039/x0xx00000x

H. R. Rachita and H. A. Nagarajaram

Received 00th January 2012,
Accepted 00th January 2012

DOI: 10.1039/x0xx00000x

www.rsc.org/

Viruses, despite having small genomes and few proteins, make array of interactions with host proteins as they solely depend on host machinery for their replication and reproduction. Hence, analysis of Human-Virus Protein-Protein Interaction Network (Hu-Vir PPI network) helps us to gain certain insights into molecular mechanisms underlying hijacking of host cell machinery by viruses for their perpetuation. Here we report an analysis of Human-Virus Bridged PPI Network (HVBN) that has led us to identify viral Articulation Points (VAPs) which connect unconnected components of Human-PPI (Hu-PPI) network. VAPs cross-link peripheral nodes to the giant component of Hu-PPI network. VAPs interact with number of relatively lower topologically central human proteins and are conserved among related viruses. The linked nodes comprise of those that are mostly expressed during viral infection as well as those that are found exclusively in some metabolic pathways indicating that these novel viral mediation of certain human protein-protein interactions may form basis for virus-specific tuning of host machinery. Functional importance of VAPs and their interaction partners in virus replication make them potential drug targets against viral infection. Our investigations also lead to the discovery of an example of Human Endogenous Retrovirus (HERV) encoded protein syncytin as an Articulation Point (AP) in Hu-PPI network suggesting that the VAPs may get retained in a genome if they result in any beneficial function to the host.

Introduction

Viruses depend on their host cells to proliferate. Viruses are metabolically inert when they are outside their host cells however; they literally take over the host's functions after infecting them by means of inserting their genetic material into the host cells. Viruses hijack host cellular system and make it produce more of viral proteins and genetic material than its usual products[1]. In order to achieve this, viral proteins make an array of interactions with the proteins in host cells. At the cellular level host functions can be represented as protein-protein interaction (PPI), regulatory and metabolic networks and some of which are hijacked by viruses for their propagation[2].

Literature abounds with studies on various aspects of individual PPI networks pertaining to systems ranging from organelles to whole organisms. However, reported studies on host-pathogen PPI networks to unravel the underlying intricacies of infection have only been a few. First systematic analysis of host and virus PPIs was reported for human and the two herpes viruses Varicella Zooster Virus (VZV) and Kaposi Sarcoma associated Herpes Virus (KSHV). This study revealed that virus PPI networks lack clustering and form highly connected cohesive modules when merged with host networks[3]. Viruses were found to adopt cellular network characteristic features when merged with host PPI networks[3,4]. Systematic analysis of Human-Virus (Hu-Vir) PPIs have revealed that the viral proteins target human proteins occupying 'hub' and 'high betweenness' which are involved in cell cycle regulation and signalling pathways[5–8]. HIV-1 targets highly intertwined nodes in human PPI network known as rich clubs[9]. Mass spectrometry approach has been utilised to identify complexes formed by HIV-1 proteins in human cell lines[10]. Interactions with bottlenecks, hubs and rich clubs allow viruses to direct most of the host machinery for synthesising viral proteins and genomes[3,6–9].

Any perturbation in the homeostasis of host caused by pathogen is actually dissipated throughout the host PPI network, hence it is imperative to study viral interactions with human host in the context of Hu-PPI network. Merging human-virus (Hu-Vir) PPI and Human PPI (Hu-PPI) network has been an important approach in understanding role of human proteins in viral pathogenesis[3,4,7]. However, utilisation of such merged PPI networks to study the network positions of viral proteins has not been explored so far. Availability of PPI data for human and 70 viruses prompted us to undertake network level analysis of human-virus system. In the present study we have merged the known Hu-Vir PPI pertaining to each of the 70 viruses with Hu-PPI network. We refer the resultant networks as human viral bridge networks (HVBNs). A HVBN is essentially a static representation of known protein-protein interactions in a typical virus infected host-cell, and hence helps in a better understanding

Molecular BioSystems Accepted Manuscript

of the PPIs underlying viral pathogenesis. While analysing the network properties of the HVBNs we observed that some of the viral proteins occupy certain nodal positions which when removed lead to an increase in the connected components of HVBNs. These viral nodes bridge (connect/cross-link) some of the unconnected components of the human PPI network. The nodes that connect unconnected components in a network have been referred to as Articulation points (APs)[11] and hence these viral proteins can be referred to as viral APs (VAPs). APs have been of interest in various other non-biological networks such as wireless sensor networks, road networks and scientific citation networks as they are the sole connectors mediating connections between two or more different components in the network[12–14]. When APs are targeted the entire network gets disrupted[12] as such action disconnects components of the network. Removal of APs is expected to cause lethal effects on biological networks[15]. It is to be noted that despite their importance, surprisingly the literature lacks sufficient coverage of APs in PPI networks. As HVBN represents all the known protein interactions underplaying the virus life cycle in human host we identified and studied VAPs in the context of functional enrichment of human proteins they connect and the pathways they cross-link. Interestingly we found that VAPs bridge human peripheral nodes to the giant component and most of the connected peripheral nodes are not represented in the known human PPIs and hence are probably expressed during viral infection. This observation made us to think that the VAPs identified in our study and their novel cross-linking functions may provide a basis for further exploratory studies on viral hijacking of the host system. We also found an example of endogenous retroviral protein forming an AP in Hu-PPI network hinting that VAPs may also get accepted during evolution if they render selective advantage to the host.

Methods and Materials

HUMAN PAIR-WISE PPI DATA: Human PPI (Hu-PPI) data were downloaded from the two curated human protein- protein interactions resources: IntAct (26974 protein pair-wise interactions)[16] and HPRD (Human Protein Reference Database) (38788 interactions)[17] and the union of these data was made as a representative dataset of human protein-protein interactions which was further curated manually and using NCBI BLAST[18] to remove ambiguous and obsolete entries. The union dataset, so formed, consists of 49772 PPIs involving 10446 human proteins and used as gold standard Hu-PPI network in the current study (Supplementary table S1).

HUMAN-VIRUS BIPARTITE PROTEIN INTERACTION DATA: Hu-Vir PPIs data were downloaded from Pathogen Interaction Gateway (PIG)[19] and VirusMINT[20]. To these datasets we added published yeast two-hybrid interactions for Vaccinia virus[21], Dengue virus[22], HTLV-1 and HTLV-2[23] with human proteins after removing identifiers pertaining to pseudo genes, non-coding mRNA and obsolete entries as these were not available in above mentioned databases.

The dataset was curated further by removing obsolete and duplicate entries. Later BLAST[18] was used to merge strain specific Hu-Vir PPIs and also to remove ambiguous human and viral protein identifiers (Supplementary Table S1). For human proteins 98% identity was used as cut-off and for viruses 30% was used since different isolates of viruses are highly variable as they undergo mutations rapidly. The non-redundant dataset comprises of 3389 unique interactions between 267 viral proteins from 70 different viruses and 1735 human proteins. Viruses belonging to all seven Baltimore classes of viruses[24] namely dsDNA, ssDNA, dsDNA-RT, (+) ssRNA, (-) ssRNA, ssRNA-RT and dsRNA are represented in the dataset (Supplementary table S2).

HUMAN-VIRUS BRIDGED PROTEIN-INTERACTION (HVBN) NETWORK: Hu-Vir Bridged Network (HVBN) for a virus is constructed by merging the known Hu-Vir PPI network of that virus with the known Hu-PPI network (please see Fig. 1 for an illustration). We constructed HVBNs for 70 different viruses and were analysed to identify viral articulation points (VAPs) using igraph[25] package in R[26]. Clusters (i.e., different connected components) of the Hu-PPI network were identified using 'clusters' function in igraph package for R. In a cluster each pair of nodes has a path between the two nodes within the same cluster. Large networks actually have one largest connected component referred to as giant component and few surrounding mini clusters with few nodes[27]. Hu-PPI network and HVBNs are no exceptions to this property. The human proteins which are not part of giant component in Hu-PPI network were considered as Peripheral Nodes. Into this list of peripheral nodes we also included those proteins with no known human interaction partners but become 'connected' in HVBN via viral proteins as a consequence of their known interaction with viral proteins. It turned out that the viral proteins connect peripheral nodes to giant component of Hu-PPI network in HVBN were found to be articulation points analogous to the articulation points (APs) in the other networks and we hereafter refer to such viral proteins as viral APs (VAPs).

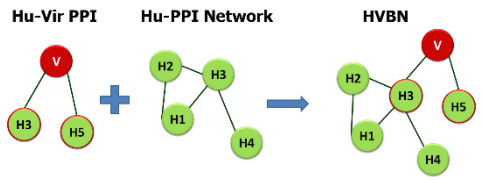


Figure 1- A schematic representation of HVBN construction. Hu-Vir PPI and Hu-PPI networks are merged to construct HVBN (red node (V)- viral protein, light green node(H1-H5)- human proteins).

RANDOMISATION OF HU-PPI NETWORK AND HVBN: In order to validate VAPs, each of the HVBNs was randomised and the frequencies with which nodes reappear as VAPs were calculated. In order to generate a randomised HVBN a two-step randomisation procedure was adopted. In the first step the bipartite Hu-Vir PPI network and the Hu-PPI network was randomised separately such that during their randomisation only edges are switched keeping node degree constant. In the second step the resultant randomised networks were merged to give to a randomised HVBN. We generated a total of 10^4 randomised

Molecular BioSystems Accepted Manuscript

This journal is © The Royal Society of Chemistry 2012

HVBNs for every virus and the frequencies of reappearances of VAPs were calculated. The igraph function 'rewire' in R was used for randomisation with iteration set to a value of 100.

CO-EXPRESSION, CO-LOCALISATION AND PATHWAY ANALYSIS OF THE HUMAN PROTEINS BRIDGED BY VAPS: Proteins have to be co-expressed and co-localised to physically interact[28–30]. We studied the co-expression and co-localisation of human proteins bridged by VAPs as such human proteins along with VAPs represent potential protein complexes. In order to determine co-expression of human proteins we made use of the gene expression data available at BioGPS[31]. These gene expression data are GC-RMA normalised NCBI GEO[32] data with GSE1133 GEO probes. This dataset consists gene expression values corresponding to 84 tissues that includes few disease conditions also. For the present analysis we excluded these disease conditions (Burkitts lymphoma, colorectal cancer and leukaemia related tissues) and used expression data from 78 tissues. We considered gene as "expressed" when its expression value is found more than the median of the whole expression data[33,34]. It has been shown that the median value cut-off yields better match with known tissue-wise expression information available in HPRD and SWISS-PROT[34]. If bridged proteins have expression value greater than median then that pair was considered to be co-expressed for further localisation and pathway analysis. If the protein pairs are co-expressed then they are given score "1" otherwise "0" (Fig. 2). Similarly for co-expressed pairs their co-localisation was checked and given a score "1" (co-localised) or "0" (not co-localised) (Fig. 2). Sub-cellular localisation information was downloaded from LOCATE database[35] which provides experimentally known localisation information for proteins curated from literature. Co-expressed and co-localised pairs were mapped on to Kyoto Encyclopedia of Genes and Genomes (KEGG)[36] pathways to find out the pathways that are cross-linked by VAPs. Later Bipartite KEGG pathway interaction network was constructed such that there is an edge between two KEGG pathways if proteins in those pathways are bridged by VAPs and that are co-expressed and co-localised. Bipartite pathway network was visualised in Cytoscape[37].

Results and Discussion

Comparison of human proteins in the curated Hu-PPIs and Hu-Vir bipartite interactions revealed that 176 human proteins interact only with viral proteins (Supplementary Fig. S1) indicating that these proteins are perhaps expressed only during viral infections. We compared functions of these proteins with functions of peripheral proteins in Hu-PPI network, using Gene Ontology (GO)[38] with the help of DAVID bioinformatics resource[39] and found that peripheral nodes in Hu-PPI network are involved in metabolic processes where as those found interacting only with viral proteins are enriched in functions related to chromatin remodelling, ion binding, stress related pathways, response to foreign stimulus, virus infectious cycle, virus reproductive processes and transcription (Supplementary Fig. S2-S4).

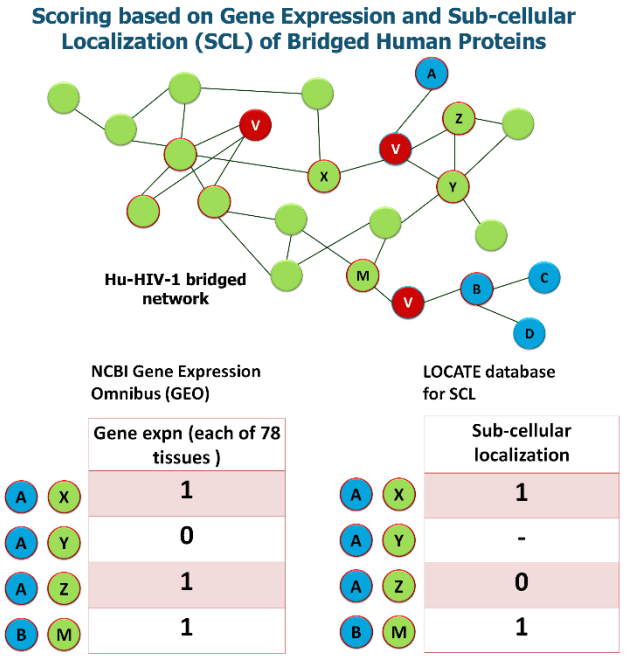


Figure 2- A schematic representation of Hu-Vir Bridged PPI Network (HVBN) with scores for co-expressed and co-localised pairs. The nodes belonging giant component are shown in light green whereas peripheral nodes are shown in blue. All possible pairs were formed between peripheral proteins (blue node with red outline) and nodes in giant component (light green with red outline) that form first neighbours of VAPs 'V' (red with blue outline). Such pairs are assigned with a score of 1 if they are co-expressed and co-localised.

Viral articulation points (VAPs) in Human Virus Bridge Network (HVBN)

Construction and analysis of HVBNs corresponding to 70 viruses revealed that only 14 HVBNs harbour VAPs. Interestingly, these 14 viruses include only ssRNA and dsDNA viruses (Table 1). The reason for not finding any of the ssDNA and dsRNA VAPs could simply be due to underrepresentation of interaction data from these viruses as ssDNA viruses are mostly bacteriophages and dsRNA viruses are mostly plant viruses; viruses infecting animals are mostly ssRNA or dsDNA viruses that are well investigated. Nonetheless, identification of VAPs in some of the HVBNs revealed a novel functional facet of some viral proteins as connectors of the unconnected human proteins. All together we found a total of 63 VAPs from 14 HVBNs.

In order to validate VAPs we calculated their frequencies in 10^4 randomised HVBN for each of the 14 viruses. We found that VAPs reappear more frequently than non-VAPs (p-value <0.001, Kolmogorov-Smirnov (KS) Test) (Fig. 3). It may be argued that the frequency of reappearance of VAPs in randomised networks is due to their promiscuous binding nature (i.e., their high degree) with human proteins. Our investigation revealed that the degree of a VAP is poorly correlated to its frequency of reappearance in a randomised HVBN (Spearman correlation r^2 =0.24, p = 0.055) indicating that their promiscuous binding nature may not be the reason for their high frequency reappearances in the randomised networks.

Molecular BioSystems Accepted Manuscript

Table 1. Details of viruses and their proteins acting as VAPs in HVBNs.

Virus	Classification	No of VAP (total no of proteins in virus)	Viral APs (VAPs)	No. of peripheral human proteins connected via viral APs
BLV*	ssRNA-RT, *Retroviridae*	1 (6)	Tax	1
Dengue Virus	(+)ssRNA, *Flaviviridae*	7 (10)	NS5, NS4A, RNA Capsid protein, poly protein, NS2A, NS3, NS1	24
EBV	dsDNA, *Herpesviridae*	12 (94)	Denaddylase, Envelope glycoprotein, EBNA6, BZLF2, BVRF1, BRRF1, BDLF2, BFLF2, BGLF2, BDLF4, EBNA2, EBNA-LP	18
HBV	dsDNA-RT, *Hepadnaviridae*	2 (7)	C protein, X protein	3
HCV	(+)ssRNA, *Flaviviridae*	1 (2)	Polyprotein	6
HIV-1	ssRNA-RT, *Retroviridae*	11 (15)	Reverse transcriptase, Rev, P17, Gag, Nef, Vif, Tat, Vpr, Env, Vpu, Gag-pol	108
HIV-2	ssRNA-RT, *Retroviridae*	2 (15)	Gag, Nef	3
HPV HR	dsDNA, *Papillomaviridae*	2 (9)	L2, E5	1
HPV LR	dsDNA, *Papillomaviridae*	2 (9)	L2	1
HTLV-1	ssRNA-RT, *Retroviridae*	4 (6)	Tax1, Rex, HBZ, Env	22
HTLV-2	ssRNA-RT, *Retroviridae*	3 (6)	Gag, Tax2, Pol	7
INFLUENZA H1N1	(-)ssRNA, *Orthomyxoviridae*	4 (9)	PB2, PB1, PA, NP	2
SV-40	dsDNA, *Polyomaviridae*	1 (7)	Large T antigen	1
Vaccinia Virus	dsDNA, *Poxviridae*	12 (223)	mRNA capping enzyme, A27, C6, K3, E3, E8, B14, VLTF3, E7, G6, K7, Inactive chemokine binding protein	22

*Non-human virus used as a model for HTLV infection where the viral proteins interact with human proteins[40].

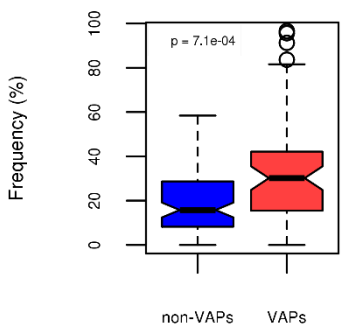


Figure 3- Frequency distribution of VAPs and non-VAPs in 10^4 randomised networks. P-value is calculated using Kolmogorov-Smirnov (KS) test.

Generally viral proteins have been found to target topologically central nodes in Hu-PPI network [3,5–9]. We, therefore, compared the topological properties of the human proteins interacting with VAPs with those interacting with non-VAPs identified from 14 different HVBNs. Our investigations revealed that the human proteins interacting with VAPs have lower centrality measures in Hu-PPI network as compared with those interacting with non-VAPs (p-value $<10^{-4}$, KS test) (Fig. 4).

Similarity among VAPs

In order to investigate how many of the VAPs are conserved among HVBNs we performed NCBI BLAST[18] searches for every VAP against the database of all the 63 VAPs and also used needle program in EMBOSS to measure the extent of pair-wise identities among 63 VAPs. We found that only the viral APs among related viruses are conserved. For example L2 protein is a conserved VAP in the HVBN of HPV-LR and HPV-HR. Tax protein of BLV, HTLV-1/2 and Nef and Tat of HIV-1/2 also share high sequence similarities.

We also investigated the extent of similarities among VAPs based on their human interaction partners as well as pathways they connect by calculating the Jaccard Similarity Index ($J(A,B) = |A \cap B|/|A \cup B|$). Interestingly, we found that VAPs, in general with high sequence similarities show low $J(A,B)$ scores, indicating that sequentially similar VAPs target uncommon proteins and pathways except HIV-1 and HIV-2 Nef. However, there were a few examples whose sequences were dissimilar but shared proteins and pathways. HTLV-1protein Tax-1 and HCV polyprotein target similar pathways as HIV-1 matrix protein, reverse transcriptase, Tat, Gag Nef, Rev, Vif and Vpu proteins (Jaccard Index >0.4). HIV-2 Vif targets similar pathways as HIV-1 nef protein (Jaccard Index>0.4). Conservation of certain viral proteins as articulation points

Molecular BioSystems Accepted Manuscript

This journal is © The Royal Society of Chemistry 2012

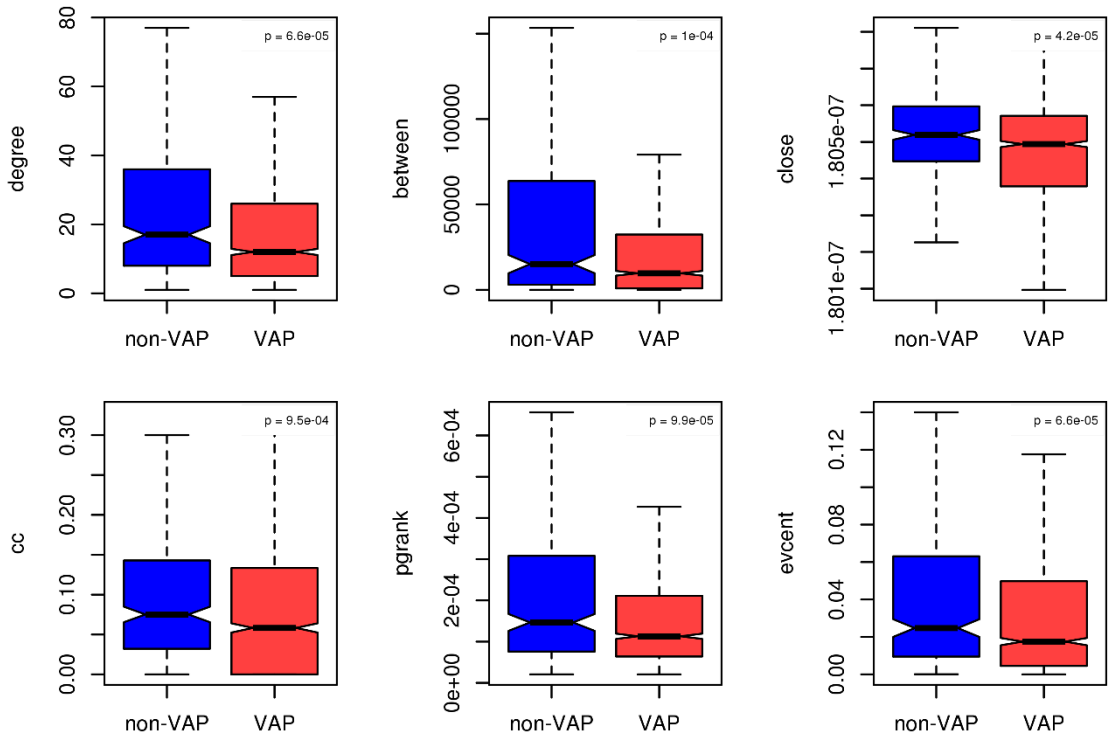


Figure 4- Network properties of human proteins interacting with VAPs* and non-VAPs. Y-axes corresponds to degree centrality (designated as degree), betweenness centrality (between), closeness centrality (close), clustering coefficient (cc), PageRank (pgrank) and Eigen vector centrality (evcent). P-values are calculated using KS test (* VAPs also connect 176 peripheral human proteins for which human PPIs are not known and hence network properties cannot be calculated for them).

indicates the central/pivotal roles of those viral proteins in the viral pathogenesis

From the network perspective viral articulation points connect the peripheral nodes to the giant component thereby increasing its size. For example, HIV-1 APs connect the largest number of peripheral proteins (104) to the human giant component followed by Dengue Virus (24 proteins), HTLV-1 (22 proteins), Vaccinia Virus (22 proteins) and Epstein Barr Virus (EBV) (18 proteins) (Table 1). The observation that most of the viruses baring HIV-1 have only few articulation points is perhaps a mere reflection of the size of the known interaction data and may not entirely be the intrinsic property of their respective interaction networks. As HIV-1 harbours the largest number of VAPs, we considered only this virus for further analyses. Annotating HIV-1 VAP interacting peripheral human proteins to genome wide disease association database using DAVID indicated that these human proteins were associated with Hepatitis C virus (HCV) and Hepatitis B virus (HBV) infection suggesting common pathways targeted by HIV-1, HCV and HBV (Fig. 5). As mentioned above VAPs of HIV-1 and HCV target similar pathways such as wnt-signalling pathways and lipid metabolism. Peripheral human proteins interacting with VAPs are also implicated in autoimmune disorders like multiple sclerosis and type II diabetes. OMIM annotation also confirmed the same result.

HIV-1 articulation points

HIV-1 codes for 15 proteins that are required at various stages of its life cycle in host[41]. Gag, Pol and Env are structural proteins in HIV-1. Env forms envelope of virus consisting of gp120 and gp41. Pol polyprotein is cleaved into the enzymes: Reverse transcriptase, RNase H, Integrase and protease that are responsible for virus genome replication and cleavage of polyproptein. Gag, a necessary protein for virus assembly and budding is synthesised as polyprotein precursor and later during virus maturation is cleaved into p17 matrix, p24 capsid, p7 nucleocapsid and p6[41,42]. HIV-1 also codes for regulatory proteins Rev and tat. Tat, a transactivation factor which is involved in altered gene expression of various host factors and viral proteins, chromatin remodelling and DNA binding[43]. HIV-1 accessory proteins Vpr, Vif, Nef and Vpu are required for nuclear transport of virus, its infectivity, and proteasome degradation of certain host factors respectively[41]. Of 15 HIV-1 proteins 11 proteins (namely Reverse transcriptase, Gag-pol, Nef, Vpu, Vpr, Rev, Matrix P17, Vif, Gag, Env, Tat) act as VAPs in HVBN (Table 2). Among HIV-1 VAPs Tat protein connects maximum number of peripheral nodes (73 nodes) followed by Gag connecting 20 peripheral nodes to the giant component. These two proteins also make large number of interactions with human proteins.

Molecular BioSystems Accepted Manuscript

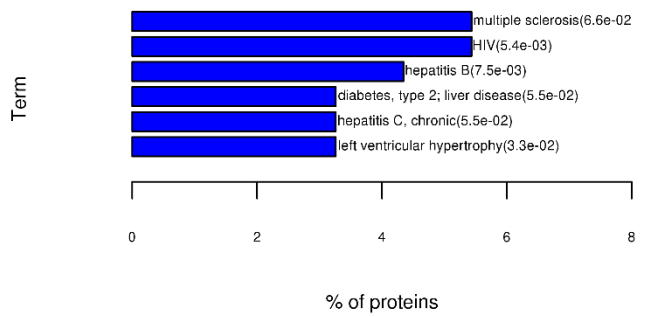


Figure 5- Disease association enrichment for the peripheral human proteins connected to giant component via HIV-1 VAPs (p-values are given within parentheses). Data was taken from Disease association database using DAVID.

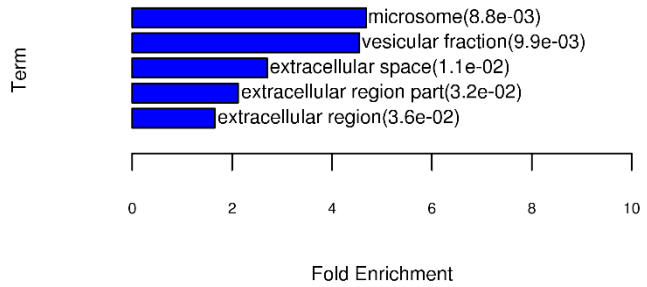


Figure 6- Enrichment of GO cellular components in human proteins connected via HIV-1 VAPs (p-values are given within parentheses) in HVBN. For each cellular component there are 4 or more human proteins.

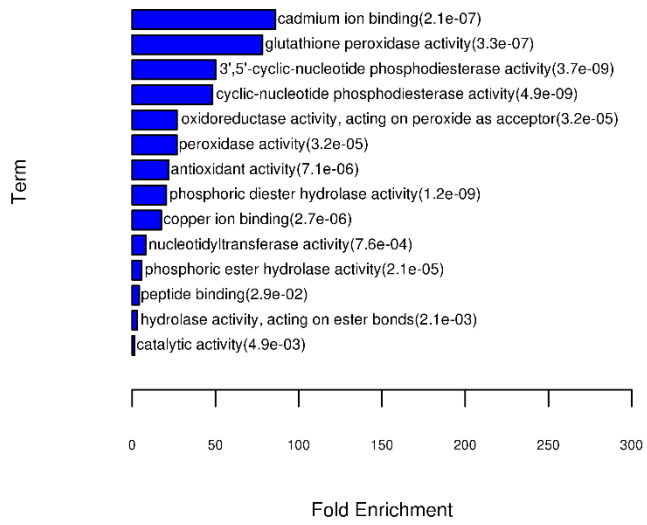


Figure 7- Enrichment of GO molecular functions in the peripheral human proteins connected via HIV-1 VAPs in HVBN (p-values are given within parentheses). For each molecular function there are 4 or more human proteins.

Functional analysis of human proteins bridged by HIV-1 VAPs

Functional analysis of peripheral human proteins connected by HIV-1 APs to the giant component was carried out using DAVID Bioinformatics database[39]. GO functional annotation of human

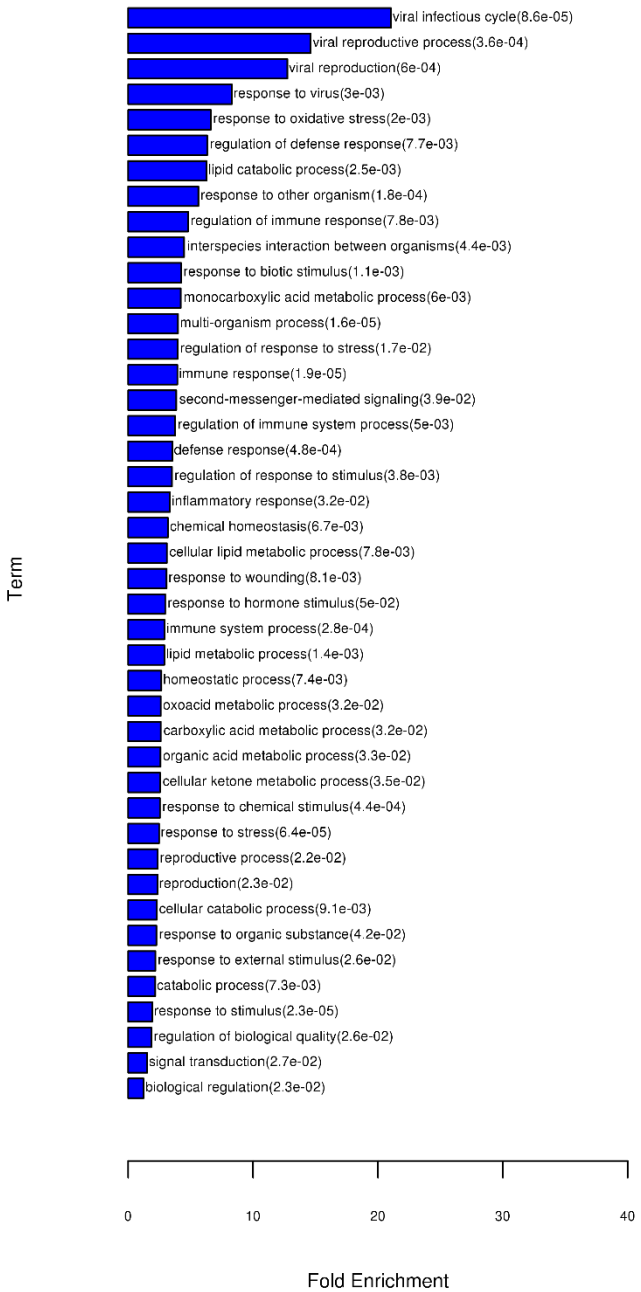


Figure 8- Enrichment of GO biological processes in peripheral human proteins connected via HIV-1 VAPs (p-values are given within parentheses). For each biological process there are at least 4 or more human proteins.

proteins connected to the main network via HIV-1 APs indicated that these proteins are either secretory or nucleosome components (Fig. 6) and bind to metal ions (Fig. 7) and are enriched in functions like metabolic pathways, stress response, interspecies interaction and response to stimulus (p-value <0.05, >3 proteins per function) (Fig. 8). It has been shown that majority

Molecular BioSystems Accepted Manuscript

This journal is © The Royal Society of Chemistry 2012

Table 2- Integration of gene expression, protein sub-cellular localisation and KEGG pathway information corresponding to pairs of peripheral and giant component proteins interacting with HIV-1 VAPs.

HIV-1 articulation point	Peripheral humans proteins	Human proteins in giant component	Pairs	Gene expression	LOCATE + GEO	LOCATE + GEO + KEGG
Reverse transcriptase	2	41	82	71	1	1
Gag-pol	3	36	108	72	0	0
Nef	3	70	210	129	4	4
Vpu	4	55	220	165	1	1
Vpr	3	77	231	148	0	0
Rev	4	86	344	252	0	0
Matrix P17	6	68	408	375	2	2
Vif	6	90	540	449	0	0
Gag	10	92	920	806	15	9
Env	21	97	2037	1340	45	15
Tat	71	552	39192	27207	270	138

of viral targets in human genome are generally involved in cell cycle regulation, regulation of apoptosis and nuclear membrane transport[6] but as we show proteins connected by VAPs are functionally distinct (Fig. 8).Viruses require metabolites during replication[44] and taking cue from our present study it can be suggested that VAPs take control over metabolic pathways.

Identifying bridges formed by HIV-1 articulation points as potential drug targets

Bridges formed by VAPs between peripheral and giant component proteins of Hu- PPI network were identified by integrating gene expression data and sub-cellular protein localisation data. All possible pairs VAP neighbours between giant component and peripheral proteins were made for each HIV-1 VAP and studied their co-expression and co-localisation. Of all 44292 possible pairs between giant component and peripheral protein pairs, 338 were found to be co-expressed and co-localised thus indicating potential bridges. Of these 170 could be mapped on to KEGG pathway annotation data (Table 2). Bipartite KEGG pathway network of co-expressed and co-localised protein bridges formed by HIV-1 VAPs was constructed such that edges connect pathways involving peripheral proteins (purple nodes) and pathways of proteins in giant component (light green) (Fig. 9). Analysis of this bipartite pathway network revealed that VAPs bridge proteins involved in lipid metabolism, tRNA biosynthesis with RNA degradation, and proteins involved in highly interconnected signalling pathways, mineral absorption with signalling and metabolic pathways (Fig. 9). This suggests that targeting HIV-1 interaction with metal absorption or RNA degradation pathways or lipid metabolism would be the key in drug development for HIV-1 infection.

The peripheral nodes which are connected by VAPs to the giant component are involved in metabolic pathways. This gives clues about how viruses take over the regulation of metabolic pathways which are essential for their replication as they provide building blocks for viral components. HIV-1 or in general any virus requires cellular metabolites along with host cellular machinery for genome replication. Lipids are required for synthesis of envelope which is made off lipid bilayer that is also rich in glycoproteins. HIV-1 connects metabolic pathways to PPI network and hence take over host metabolic processes for its replication. HIV-1 VAP interacting peripheral proteins are also associated with diabetes (glucose metabolism) and liver disease (lipid metabolism) (Fig. 5). During acute infection glucose intolerance, changes in serum lipoprotein pattern and insulin insensitivity have been observed in patients[45,46] suggesting regulation of concerned pathways by the virus. Perturbation of metabolic pathways by viruses has been shown to be associated with neuronal dementia[47] and non-alcoholic fatty liver disease[48,49]. VAPs have higher degree ($p<10^{-12}$, KS Test) in bipartite network as compared to non-VAPs. On the contrary human interaction partners of VAPs have lower centrality measures compared to those of non-VAPs. Targeting VAPs and interacting human proteins has dual advantage as it has been shown that targeted deletion of high degree nodes result in network breakdown[4,50]. Viruses interact with topologically central proteins such as hubs, bottlenecks, rich clubs and those that are involved in important processes of cell[3–7,9] but along with that they also form APs in Hu-PPI network so that they can bring together some of the central functions and hijack cellular systems. Hubs or high betweenness proteins of human PPI network definitely represent important biological functions in the host and hence have disadvantage of being drug targets for viral infection. It has been recently reported that drugs targeting highly central proteins have more side effects[51,52]. In this light, the human proteins interacting with VAPs form attractive possibility for anti-viral drug targets. Interestingly, in literature we found two such studies which revealed that inhibition of ceramide / sphingolipid pathway in HIV-1 reduced virus infectivity[53] and in HCV, it attenuated the virus and replication was strongly repressed[54].

Functional consequences of presence/absence of VAPs: Examples from literature

There are reports where the authors have demonstrated the functional importance of some of the viral proteins which we have identified as VAPs (vpu, vpr and matrix from HIV-1; rex from HTLV-1 and NS4A from DENV). As these viral proteins are VAPs in the human-viral protein-protein interaction bridge

Molecular BioSystems Accepted Manuscript

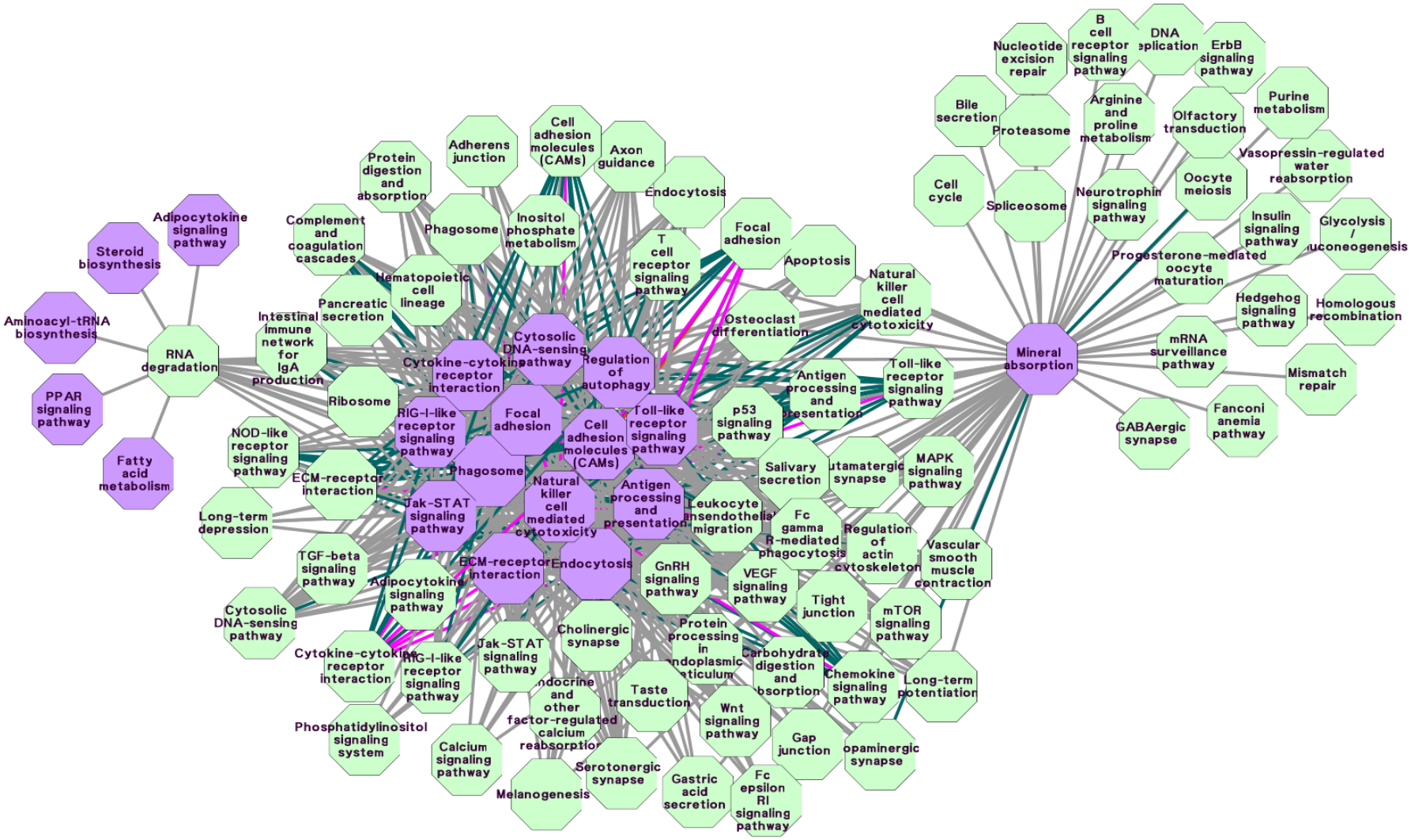


Figure 9- Bipartite KEGG pathway network of proteins bridged via HIV-1 VAPs. Pathways involving peripheral nodes are represented by purple nodes and pathways involving proteins in giant component are shown in light green nodes. Edge colour represents number of bridges between pathways; grey (1-4), green (5-9), violet (10-20) and red (>20). Network was visualised in Cytoscape[37].

Molecular BioSystems Accepted Manuscript

This journal is © The Royal Society of Chemistry 2012

networks, we believe that the observed functional consequences of their presence/absence are due to their role as VAPs. The following paragraphs give the details of those VAPs.

HTLV-1 rex protein has been shown to interfere with host nonsense mediated mRNA decay (NMD) and mRNA metabolism in HTLV-1 provirus (gag-pol mRNA) transformed HeLa cell line when compared with uninfected cell line[55]. Absence of rex compromised viral RNA stability however siRNA suppression of NMD restored the viral RNA stability[55].

Dengue virus (DENV) induced autophagy and β-oxidation of triglycerides in autophagosome to generate ATP are required for viral replication[56]. siRNA knockdown of autophagosome formation or use of β-oxidation pathway antagonist etomoxir decreased DENV RNA in infected cell lines[56]. DENV NS4A has been shown to be essential for inducing membrane curvature[57] and autophagy[58] which are linked to each other[59]. Absence NS4A failed to form of autophagosome which in turn hampered viral replication[58]. Deletion of vpu in HIV-1 has been shown to affect virion release from the infected cells suggesting its crucial role in HIV-1 lifecycle.

HIV-1 tat[46] and vpr[60] have been reported to be responsible for metabolic imbalance. In a recent study, vpr has been linked to adipose dysfunction via disruption of PPAR/GR signalling pathways in transgenic mouse model expressing HIV-1 vpr as compared with wild type mouse without vpr[60]. Presence of vpr increased lipolysis in adipocytes and resulted in hepatoosteosis in mice.

In addition to the above mentioned case studies, further exploratory studies can be conducted on other VAPs. For example, our studies have revealed that HIV-1 matrix protein bridges mitochondrial proteins CYP27B1, HSPA9, mitochondrial HARS and cAMP dependent protein kinase α and β subunits. We also know from literature that: (a) HIV-1 infection disrupts mitochondrial functions[49]; (b) CYP27B1 involvement in the activation of glycolysis along with its other functions[61] and (c) certain mitochondrial proteins are sequestered into virus during packaging[49]. These suggest that HIV-1 matrix protein can be an attractive target for further experimental studies to get more insights into mitochondrial dysfunction associated with HIV-1 infection. We, therefore, suggest further explorative studies on VAPs and their interactions in Hu-PPI network as novel drug targets to suppress viral actions in human cells.

Human Endogenous Retroviruses (HERV) proteins as articulation points in Hu-PPI network

Identification of VAPs in HVBNs led us to wonder whether viral pathogens contribute to gain of novel functions in the hosts they infect and are such novel connections beneficial to host or not. If it is so then some evidences can be seen in the human genome. It is known that the human genome has remnants of ancestral viral infections where viral genome has been incorporated through germline DNA known as human endogenous retroviruses (HERVs). HERVs form ~8% of human genome[62]. We investigated role of HERV proteins in Hu-PPI network. We took 576 human proteins which have been annotated as HERV proteins in UniProt/SwissProt[63]. Of these, we found HERV-W 7q21.2 provirus encoded protein syncytin as an AP in Hu-PPI network. Syncytin protein is required for trophoblast fusion and hence for placenta development in mammals[64,65]. The retrovirus encoded syncytin homolog is present in all placental mammals and in avian reticuloendotheliosis virus[66]. Syncytin protein which is acquired by mammals through viral infection plays very important role in evolution of mammals[65]. HERV protein syncytin acting as AP in Hu-PPI network indicates that VAPs might result in gain of function and increased complexity of the network by connecting previously unconnected components. Example of HERV-W 7q21.2 proviral protein syncytin implies that there are viral proteins which have been incorporated into the host genome and are playing very important role. It is also known that many HERV proteins are also known to be associated with autoimmune disorders[67–69] suggesting they may not be always beneficial and have their own implications in host.

Conclusions

When virus infects host cell it does not merely make interactions with host proteins but also adapts to host network topology and make novel connections in Hu-PPI network. Some of the viral proteins function as articulation points in the bridged Hu-PPI network. We show that some of the VAPs are conserved across closely related viruses. The viral articulation points connect peripheral nodes corresponding to the proteins involved in metabolic pathways or those expressed during viral infection to the giant component. Viruses along with host cellular machinery for replication also need metabolites and hence they connect metabolic pathways and bring about altered regulation of these metabolic pathways. Probing the VAP mediated bridge interactions might help in identifying novel drug targets for HIV-1. We found some examples in the literature where blocking certain host metabolic pathways attenuated viral replication[53,54]. Furthermore, there are studies conducted on viral proteins which we have identified as VAPs and these studies demonstrate the functional consequence of presence/absence of these viral proteins in relation to viral replication and infection. We propose identifying novel connections made by viral proteins as an intelligent approach towards identifying drug targets to suppress viral replication which then can be validated experimentally.

Acknowledgements

Authors thank CDFD for providing computational facilities. This work was supported by a core grant from CDFD to H.A.N. R.H.R., a registered PhD student of Manipal University, gratefully acknowledges CSIR for Senior Research Fellowship. Last but not the least, the authors thank all the anonymous reviewers for their constructive comments.

Notes

[a] Centre for DNA Fingerprinting and Diagnostics, Gruhakalpa, 5-4-399/B, Nampally, Hyderabad 500001, India. E-mail: han@cdfd.org.in; Tel: +91 40-24749367

Molecular BioSystems Accepted Manuscript

Electronic Supplementary Information (ESI) available: Supplementary Information is available as separate document. See DOI: 10.1039/b000000x/

References

1. D. Walsh and I. Mohr, *Nat. Rev. Microbiol.*, 2011, **9**, 860–75.
2. C. V Forst, *Drug Discov. Today*, 2006, **11**, 220–7.
3. P. Uetz, Y.-A. Dong, C. Zeretzke, C. Atzler, A. Baiker, B. Berger, S. V Rajagopala, M. Roupelieva, D. Rose, E. Fossum, and J. Haas, *Science*, 2006, **311**, 239–42.
4. L. Meyniel-Schicklin, B. de Chassey, P. André, and V. Lotteau, *Mol. Cell. Proteomics*, 2012, **11**, M111.014738.
5. M. A. Calderwood, K. Venkatesan, L. Xing, M. R. Chase, A. Vazquez, A. M. Holthaus, A. E. Ewence, N. Li, T. Hirozane-Kishikawa, D. E. Hill, M. Vidal, E. Kieff, and E. Johannsen, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 7606–11.
6. M. D. Dyer, T. M. Murali, and B. W. Sobral, *PLoS Pathog.*, 2008, **4**, e32.
7. V. Navratil, B. de Chassey, C. R. Combe, and V. Lotteau, *BMC Syst. Biol.*, 2011, **5**, 13.
8. R. R. Halehalli and H. A. Nagarajaram, *Unpubl. data*.
9. S. Wuchty, G. Siwo, and M. T. Ferdig, *PLoS One*, 2010, **5**, e11796.
10. S. Jäger, P. Cimermancic, N. Gulbahce, J. R. Johnson, K. E. McGovern, S. C. Clarke, M. Shales, G. Mercenne, L. Pache, K. Li, H. Hernandez, G. M. Jang, S. L. Roth, E. Akiva, J. Marlett, M. Stephens, I. D'Orso, J. Fernandes, M. Fahey, C. Mahon, A. J. O'Donoghue, A. Todorovic, J. H. Morris, D. A. Maltby, T. Alber, G. Cagney, F. D. Bushman, J. A. Young, S. K. Chanda, W. I. Sundquist, T. Kortemme, R. D. Hernandez, C. S. Craik, A. Burlingame, A. Sali, A. D. Frankel, and N. J. Krogan, *Nature*, 2012, **481**, 365–70.
11. R. E. Tarjan, *Inf. Process. Lett.*, 1974, **2**, 160–161.
12. H. H. Benahmed Khelifa, *Comput. Commun.*, 2009, **2**, 507 – 512.
13. F. Xie and D. Levinson, *Geogr. Anal.*, 2007, **39**, 336–356.
14. L. Leydesdorff, *J. Doc.*, 2004, **60**, 371–427.
15. N. Przulj, D. A. Wigle, and I. Jurisica, *Bioinformatics*, 2004, **20**, 340–8.
16. B. Aranda, P. Achuthan, Y. Alam-Faruque, I. Armean, A. Bridge, C. Derow, M. Feuermann, A. T. Ghanbarian, S. Kerrien, J. Khadake, J. Kerssemakers, C. Leroy, M. Menden, M. Michaut, L. Montecchi-Palazzi, S. N. Neuhauser, S. Orchard, V. Perreau, B. Roechert, K. van Eijk, and H. Hermjakob, *Nucleic Acids Res.*, 2010, **38**, D525–31.
17. T. S. Keshava Prasad, R. Goel, K. Kandasamy, S. Keerthikumar, S. Kumar, S. Mathivanan, D. Telikicherla, R. Raju, B. Shafreen, A. Venugopal, L. Balakrishnan, A. Marimuthu, S. Banerjee, D. S. Somanathan, A. Sebastian, S. Rani, S. Ray, C. J. Harrys Kishore, S. Kanth, M. Ahmed, M. K. Kashyap, R. Mohmood, Y. L. Ramachandra, V. Krishna, B. A. Rahiman, S. Mohan, P. Ranganathan, S. Ramabadran, R. Chaerkady, and A. Pandey, *Nucleic Acids Res.*, 2009, **37**, D767–72.
18. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, *J. Mol. Biol.*, 1990, **215**, 403–410.
19. T. Driscoll, M. D. Dyer, T. M. Murali, and B. W. Sobral, *Nucleic Acids Res.*, 2009, **37**, D647–50.
20. A. Chatr-aryamontri, A. Ceol, D. Peluso, A. Nardozza, S. Panni, F. Sacco, M. Tinti, A. Smolyar, L. Castagnoli, M. Vidal, M. E. Cusick, and G. Cesareni, *Nucleic Acids Res.*, 2009, **37**, D669–73.
21. L. Zhang, N. Y. Villa, M. M. Rahman, S. Smallwood, D. Shattuck, C. Neff, M. Dufford, J. S. Lanchbury, J. Labaer, and G. McFadden, *J. Proteome Res.*, 2009, **8**, 4311–8.
22. S. Khadka, A. D. Vangeloff, C. Zhang, P. Siddavatam, N. S. Heaton, L. Wang, R. Sengupta, S. Sahasrabudhe, G. Randall, M. Gribskov, R. J. Kuhn, R. Perera, and D. J. LaCount, *Mol. Cell. Proteomics*, 2011, **10**, M111.012187.
23. N. Simonis, J.-F. Rual, I. Lemmens, M. Boxus, T. Hirozane-Kishikawa, J.-S. Gatot, A. Dricot, T. Hao, D. Vertommen, S. Legros, S. Daakour, N. Klitgord, M. Martin, J.-F. Willaert, F. Dequiedt, V. Navratil, M. E. Cusick, A. Burny, C. Van Lint, D. E. Hill, J. Tavernier, R. Kettmann, M. Vidal, and J.-C. Twizere, *Retrovirology*, 2012, **9**, 26.
24. D. Baltimore, *Bacteriol. Rev.*, 1971, **35**, 235–41.
25. G. Csardi and T. Nepusz, *InterJournal*, 2006, **Complex Sy**.
26. R. R Development Core Team, *R Found. Stat. Comput.*, 2011, 1, 409.
27. M. E. J. Newman, *Proc. Natl. Acad. Sci.*, 2001, **98**, 404–409.
28. B. A. Shoemaker and A. R. Panchenko, *PLoS Comput. Biol.*, 2007, **3**, e42.
29. B. A. Shoemaker and A. R. Panchenko, *PLoS Comput. Biol.*, 2007, **3**, e43.
30. H. B. Fraser, A. E. Hirsh, D. P. Wall, and M. B. Eisen, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 9033–8.
31. C. Wu, C. Orozco, J. Boyer, M. Leglise, J. Goodale, S. Batalov, C. L. Hodge, J. Haase, J. Janes, J. W. Huss, and A. I. Su, *Genome Biol.*, 2009, **10**, R130.
32. T. Barrett, D. B. Troup, S. E. Wilhite, P. Ledoux, C. Evangelista, I. F. Kim, M. Tomashevsky, K. A. Marshall, K. H. Phillippy, P. M. Sherman, R. N. Muertter, M. Holko, O. Ayanbule, A. Yefanov, and A. Soboleva, *Nucleic Acids Res.*, 2011, **39**, D1005–10.
33. A. I. Su, M. P. Cooke, K. A. Ching, Y. Hakak, J. R. Walker, T. Wiltshire, A. P. Orth, R. G. Vega, L. M. Sapinoso, A. Moqrich, A. Patapoutian, G. M. Hampton, P. G. Schultz, and J. B. Hogenesch, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 4465–70.
34. M. Kiran and H. A. Nagarajaram, *J. Proteome Res.*, 2013.
35. J. Sprenger, J. Lynn Fink, S. Karunaratne, K. Hanson, N. A. Hamilton, and R. D. Teasdale, *Nucleic Acids Res.*, 2008, **36**, D230–3.
36. M. Kanehisa, *Nucleic Acids Res.*, 2000, **28**, 27–30.
37. P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker, *Genome Res.*, 2003, **13**, 2498–504.
38. M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock, *Nat. Genet.*, 2000, **25**, 25–9.
39. D. W. Huang, B. T. Sherman, and R. A. Lempicki, *Nat. Protoc.*, 2009, **4**, 44–57.
40. H. El Hajj, R. Nasr, Y. Kfoury, Z. Dassouki, R. Nasser, G. Kchour, O. Hermine, H. de Thé, and A. Bazarbachi, *Front. Microbiol.*, 2012, **3**, 333.
41. A. D. Frankel and J. A. Young, *Annu. Rev. Biochem.*, 1998, **67**, 1–25.
42. E. O. Freed, *Virology*, 1998, 251, 1–15.

Molecular BioSystems Accepted Manuscript

 This journal is © The Royal Society of Chemistry 2012

43. B. Romani, S. Engelbrecht, and R. H. Glashoff, *J. Gen. Virol.*, 2010, **91**, 1–12.
44. E. W. Birch, N. A. Ruggero, and M. W. Covert, *PLoS Comput. Biol.*, 2012, **8**, e1002746.
45. K. Sammalkorpi, *J. Intern. Med.*, 1989, **225**, 15–9.
46. K. Sammalkorpi, V. Valtonen, Y. Kerttula, E. Nikkilä, and M.-R. Taskinen, *Metabolism*, 1988, **37**, 859–865.
47. Y. Liu, M. Jones, C. M. Hingtgen, G. Bu, N. Laribee, R. E. Tanzi, R. D. Moir, A. Nath, and J. J. He, *Nat. Med.*, 2000, **6**, 1380–7.
48. E. M. Brunt, *Nat. Rev. Gastroenterol. Hepatol.*, 2010, **7**, 195–203.
49. P. Pérez-Matute, L. Pérez-Martínez, J. R. Blanco, and J. A. Oteo, *Oxid. Med. Cell. Longev.*, 2013, **2013**, 493413.
50. R. Albert, H. Jeong, and A. Barabasi, *Nature*, 2000, **406**, 378–82.
51. H. W. Han, J. H. Ohn, J. Moon, and J. H. Kim, *Nucleic Acids Res.*, 2013, **41**, 9209–17.
52. X. Wang, B. Thijssen, and H. Yu, *PLoS Comput. Biol.*, 2013, **9**, e1003119.
53. N. Izquierdo-Useros, M. Naranjo-Gómez, J. Archer, S. C. Hatch, I. Erkizia, J. Blanco, F. E. Borràs, M. C. Puertas, J. H. Connor, M. T. Fernández-Figueras, L. Moore, B. Clotet, S. Gummuluru, and J. Martinez-Picado, *Blood*, 2009, **113**, 2732–41.
54. F. Amemiya, S. Maekawa, Y. Itakura, A. Kanayama, A. Matsui, S. Takano, T. Yamaguchi, J. Itakura, T. Kitamura, T. Inoue, M. Sakamoto, K. Yamauchi, S. Okada, A. Yamashita, N. Sakamoto, M. Itoh, and N. Enomoto, *J. Infect. Dis.*, 2008, **197**, 361–70.
55. K. Nakano, T. Ando, M. Yamagishi, K. Yokoyama, T. Ishida, T. Ohsugi, Y. Tanaka, D. W. Brighty, and T. Watanabe, *Microbes Infect.*, 2013, **15**, 491–505.
56. N. S. Heaton and G. Randall, *Cell Host Microbe*, 2010, **8**, 422–32.
57. S. Miller, S. Kastner, J. Krijnse-Locker, S. Bühler, and R. Bartenschlager, *J. Biol. Chem.*, 2007, **282**, 8873–82.
58. J. E. McLean, A. Wudzinska, E. Datan, D. Quaglino, and Z. Zakeri, *J. Biol. Chem.*, 2011, **286**, 22147–59.
59. W. Fan, A. Nassiri, and Q. Zhong, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 7769–74.
60. N. Agarwal, D. Iyer, S. G. Patel, R. V Sekhar, T. M. Phillips, U. Schubert, T. Oplt, E. D. Buras, S. L. Samson, J. Couturier, D. E. Lewis, M. C. Rodriguez-Barradas, F. Jahoor, T. Kino, J. B. Kopp, and A. Balasubramanyam, *Sci. Transl. Med.*, 2013, **5**, 213ra164.
61. J. G. J. Hoenderop, H. Chon, D. Gkika, H. A. R. Bluyssen, F. C. P. Holstege, R. St-Arnaud, B. Braam, and R. J. M. Bindels, *Kidney Int.*, 2004, **65**, 531–9.
62. E. S. Lander, L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C. Worley, C. M. Rives, J. H. Gorrell, M. L. Metzker, S. L. Naylor, R. S. Kucherlapati, D. L. Nelson, G. M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, D. R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H. M. Lee, J. Dubois, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. R. Cox, M. V Olson, R. Kaul, N. Shimizu, K. Kawasaki, S. Minoshima, G. A. Evans, M. Athanasiou, R. Schultz, B. A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W. R. McCombie, M. de la Bastide, N. Dedhia, H. Blöcker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D. G. Brown, C. B. Burge, L. Cerutti, H. C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E. Eichler, T. S. Furey, J. Galagan, J. G. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kaspryzk, S. Kennedy, W. J. Kent, P. Kitts, E. V Koonin, I. Korf, D. Kulp, D. Lancet, T. M. Lowe, A. McLysaght, T. Mikkelsen, J. V Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. Schultz, G. Slater, A. F. Smit, E. Stupka, J. Szustakowski, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, K. H. Wolfe, S. P. Yang, R. F. Yeh, F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand, A. Patrinos, M. J. Morgan, P. de Jong, J. J. Catanese, K. Osoegawa, H. Shizuya, S. Choi, Y. J. Chen, and J. Szustakowki, *Nature*, 2001, **409**, 860–921.
63. R. Apweiler, A. Bairoch, and C. H. Wu, *Curr. Opin. Chem. Biol.*, 2004, **8**, 76–80.
64. A. Dupressoir, C. Vernochet, O. Bawa, F. Harper, G. Pierron, P. Opolon, and T. Heidmann, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 12127–32.
65. J. R. Harris, *FEBS Lett.*, 1991, **295**, 3–4.
66. R. Gifford and M. Tristem, *Virus Genes*, 2003, **26**, 291–315.
67. F. P. Ryan, *Curr. Neuropharmacol.*, 2011, **9**, 360–9.
68. C. Serra, G. Mameli, G. Arru, S. Sotgiu, G. Rosati, and A. Dolei, 2009.
69. G. Mameli, V. Astone, G. Arru, S. Marconi, L. Lovato, C. Serra, S. Sotgiu, B. Bonetti, and A. Dolei, *J. Gen. Virol.*, 2007, **88**, 264–74.

Molecular BioSystems Accepted Manuscript