**Virus-host protein-protein interactions in human disease**

Norman Goodacre1, Prajwal Devkota2, Eunhae Bae1, Stefan Wuchty2,3,4,5\*, Peter Uetz6\*

1 Division of Viral Products, Office of Vaccines Research and Review, Center for Biologics Evaluation and Research, U.S. Food and Drug Administration, Silver Spring, Maryland, USA

2 Dept. of Computer Science, Univ. of Miami, Coral Gables, FL 33146, USA

3 Center for Computational Science, Univ. of Miami, Coral Gables, FL 33146, USA

4 Dept. of Biology, Univ. of Miami, Coral Gables, FL 33146, USA

5 Sylvester Comprehensive Cancer Center, Miller School of Medicine, University of Miami, Miami, FL 33136, USA

6 Center for the Study of Biological Complexity, Virginia Commonwealth University, Richmond, VA 23284, USA

\* Corresponding authors: [wuchtys@cs.miami.edu](mailto:wuchtys@cs.miami.edu), [peter@uetz.us](mailto:peter@uetz.us)

**Abstract**

Viruses infect their hosts by a series of protein-protein interactions, starting with virus binding to surface receptors, and concluding with virus assemby and egress of complete virus particle. We discuss the coverage of human viruses and how well their proteins have been studied for host-virus interactions: only 4 virus families have accumulated more than 1000 PPIs when all their PPIs are combined, namely Orthomyxoviruses (5,494), herpesviruses (5,423), papillomaviruses (3,927) and retroviruses (2,285). Thus, some viruses have been extremely well studied, with some viruses, such as HIV, having more than a hundred interactions identified for each of its proteins. It remains unclear how many of these interactions are physiologically relevant. However, for many mecially important viruses there are numerous genome sequences available but only few interactions are known (e.g. rhinoviruses). We discuss the conclusions that can be drawn from large- and small-scale PPI studies in human viruses, how they reflect the relevance of important viruses for human health, and finally compare them to some insights gained from phage-bacterial interactions.

**Introduction**

Viruses are, together with bacteria, the most important pathogens on earth. While most bacteria are still easy to treat with antibiotics, viruses are much harder to control. That is a direct consequence of their nature, being composed of only a few nucleic acids and proteins and sometimes lipids and a few other compounds. Given their simple composition and structure viruses are completely dependend on their hosts, and hence they have to extensively interact with host proteins and other host components.

In this review we provide an overview of virus diversity and how it related to the diversity of host-virus diversity. We surmise that medically important viruses have received more attention and thus more research, hence more interactions should be known and understood. While this is often the case, there is no simple correlation. In fact, being highly pathogenic does not mean a virus is easy to study or it may not be of interest, given a very restricted geographic range or very narrow host range. For some viruses, such as Zika, the threat is relatively recent so that research has only ramped up during the past few years and there is no extensive data available yet.

**Diversity of human viruses**

Most people are infected by one or more viruses. Wylie et al. detected an average of 5.5 viral genera in each of 102 healthy individuals. Given that no more than 5 habitats were screened (nose, skin, mouth, vagina, and stool) we can safely assume that most people have dozens of different viruses in their body. However, only few of them lead to clinical symptoms or disease. We have thus compiled data on the diversity of human viruses by looking both at sequence diversity (**Table 1**) and epidemiology (**Table 2**).

**Virus-host interactions in humans**

Our level of understanding human-virus PPIs is highly biased towards a few well-studied viruses. For instance, only 2 viruses have more than 1000 interactions listed in VirusMentha, namely Epstein-Barr Virus (EBV) with 1766 and HIV-1 with 1304 PPIs, respectively (**Table X**). Only 4 virus families have more than 1000 PPIs when all their PPIs are combined, namely Orthomyxoviruses (5494), herpesviruses (5423), papillomaviruses (3927) and retroviruses (2285), respectively. While these numbers are roughly similar, they represent vastly different genome sizes and virus diversity within families. For instance, HIV encodes only about 10 proteins while it has more than a 100 interactions per protein. EBV, by comparison, encodes about 85 proteins and thus each protein interacts with “only” 20 proteins on average.

**How realiable are published virus-host interactions?**

There are very few studies that systematically validated human-virus interactions for their biochemical or even physiological validity. Blasche et al. found 60 “high-confidence” interactions among phage lambda and its host, E. coli, while only 30 interactions had been found during intense study over 50 years. Only X out of 60 had been found before, hence we can infer a false negative rate of X and a substantial false positive rate.

Among the first attempts to validate human virus-host interactions was our study of KSHV-human interactions (1). We predicted homologous interactions from experimental data *in S. cerevisiae, Caenorhabditis elegans*, and *Drosophila melanogaster*. That is, we predicted interactions among KSHV and human proteins when either *S. cerevisiae, C. elegan*s, or *D. melanogaster* had interacting orthologs. Although this is a rather far-fetched approach, we predicted 20 interactions between 8 KSHV and 20 human proteins. Nineteen of these 20 virus-host interactions were tested by CoIP and an unexpectedly large percentage (13 out of 19, or 68%) could be confirmed.

Zhang et al. found 109 interactions among 33 Vaccinia and X human proteins, of which 27 were tested by GST pull-downs (2). This translates to a 63% validation rate. While these numbers appear to be rather high, one has to remember that certain subsets were selected and thus do not necessarily represent an unbiased validation rate of complete Y2H data sets.

However, systematic Y2H screens have long suggested that there may be many legitimate and reproducible interactions that nevertheless have no physiological function – or at least minor functions so that they are difficult to detect (Uetz et al, unpublished).

In fact, interactions can and most likely do evolve relatively quickly, especially among virus-host interactions, so that a substantial number of bona fide interactions may arise that either have only minor advantages or simply do not have any disadvantage to a cell.

**How many interactions does a virus require?**

The sheer number of host-virus interactions that have been found for many viruses may suggest that we have identified all interactions. This may be close to true but these numbers are also so large that they are likely to contain a large number of false positives. So, the key question is: how many interactions does a virus need or use to infect an organism and how do we identify the physiological interactions among those tha have been found overall?

These are two different questions and both are hard to answer. Bacteriophage may serve as a model: 50 years of research have identified about 30 host-virus interactions between *E. coli* and phage lambda, which encode ~4000 and 73 proteins, respectively (3). Let’s assume that most host-virus interactions in this system have been identified. A large-scale analysis of *E. coli*-lambda interactions revealed 62 high-confidence set (among a raw data set of 631 PPIs total) (4). However, among the 62 high-confidene PPIs only two were previously known to be physiological, but the role of the other 60 remains unknown. It is possible that lambda is unusual in the sense that many proteins are processed during maturation and thus interactions are more difficult to detect.

Protein processing seems to be less common among phage T7 and its 55 proteins are known to be involved in only 15 interactions with its host (3). Both T7 and lambda have about 30-40 interactions among virus proteins, which are obviously easier to detect but possibly more common, given the more elaborate virus structure in these tailed phage when compared to the often simple-structured human viruses.

Unfortunately, we could not find any published comprehensive analysis or even a review of published interactions among human and virus proteins that attempted to evaluate these interactions for their plausibility or physiological relevance.

**Human targets of virus proteins**

During the last decade numerous protein interaction interfaces between human viruses and their host cells have been investigated. Mostly such efforts focused on, Hepatitis C virus (5-8), Human Immunodeficiency Virus (9, 10), Influenza A virus (11), Herpes (12) and other viruses including Epstein-Barr (13, 14) and Dengue (15) and numerous other viruses (16). As a consequence, many different databases have been designed to capture the abundance of such interactions (**Table 1**). Currently, a number of Web-based resources aim to integrate pathogen–host molecular interactions and related data available in the literature. Some of them store data on only one specific pathogen species such as HCVpro (17) and HIV-1 Human Interaction Database @NCBI (18). The resources based on a wider range of human specific viruses are VirHostNet (19), VirusMentha (20), PHIDIAS (21), HPIDB (22), and PHISTO (23) are viral databases that include interactions between human host and viral proteins.

In particular, HCVPro (HCV interaction database) is dedicated to only HCV, cataloging the characterized protein interactions for intraviral and virus–human systems. Additionally, it includes information on the structure and functions of HCV proteins (17). The HIV-1 Human Protein Interaction Database @NCBI includes the interactions between HIV-1 and human proteins where the majority of the protein interaction data are indirect (e.g., upregulation, modification) while the rest are directed interactions that relate to physical interactions (18). In turn, databases developed specifically for virus host protein interactions such as VirHostNet (19), VirusMentha (20) and HPIDB (22). VirHostNet (Virus–Host Network) is one of the earliest PHI resources specialized in the management and analysis of integrated virus–virus, virus–human host, and human host–host protein interaction networks coupled to their functional annotations. The recently developed tool, VirusMentha is another virus-virus and virus–host protein interaction resource and the most comprehensive viral PHI data source without limitation with respect to virus species or host organisms. HPIDB (Host–Pathogen Interaction Database) is not limited to any pathogen or host regarding pathogen–host PPI data.

Finally, the Web-based PHI databases comprising all pathogen types with known interactions are PHIDIAS (23) and PHISTO (23) . PHIDIAS (Pathogen–Host Interaction Data Integration and Analysis System) stores data on genome sequences, conserved domains, and gene expression data related to PHIs. In addition to data storage, PHIDIAS offers the analysis of these data. PHISTO (Pathogen-Host Interaction Search Tool) is a comprehensive PHI database including data for all pathogenic microorganisms for which experimental protein interactions with human are available.

Mostly, human host-virus protein interaction data in the above PHI databases are integrated mainly from other PPI databases using automatic integration tools such as PSICQUIC (24) and by manual curation from the literature. While the previously mentioned databases focused on host-virus interactions. However, commonly used proteoin-protein databases collect interactions between human host and virus proteins as well, including databases such as BIND (25), BioGrid (26), DIP (27), HPRD (28), IntAct (29), iRefIndex (30), MINT (30) and Reactome (31).

Given the abundance of human host-viral interactions we collected such information from the xxx databases (**Table 2**). Notably, we collected 5,495 interactions with human host proteins that involving proteins of orthomyxoviridae. Such a set of interactions is mostly dominated by interactions that occurred between proteins of the Influenza A virus. Similarly, we found 5,423 interactions of herpesviridea, where most interaction were experimentally determined using the Herpes simplex virus. While our set provided 3,927 interactions of papillomaviridae where most interactions were found with the human papilloma virus, 2,285 interactions were mostly provided by the HIV-1 virus. Pooling all remaining interactions we obtained 1,193 interactions. In **Fig. 1**, we observed that the sets of human proteins that were targeted by the different virus families substantially overlapped. The abundance of such virus-host interaction data prompted topological analysis of networks thus obtained. Using a network of more than 50,000 interactions between more than 8,000 human proteins from the HINT database (32), we determined the enrichment of viral targets as a function of their degree. In **Fig. 2A**, we found that targets of all families preferably were enriched in increasingly interacting human proteins. Such a result corroborates previous generally made observations that viral targets are preferably hubs (13, 33-38). As degree is a local measure of centrality, we consider betweenness centrality as a more global centrality measure. Defining the top 20% of proteins in the human protein interaction network as bottleneck nodes we determined their number of bottlenecks that were targeted by viruses of different families. Randomly sampling sets of bottleneck proteins, we observed that the expected number of targeted bottlenecks was statistically significant (**Fig. 2B**, P < 10-4) (13, 33-38). As a corollary, we counted the number of times that a protein appears in different pathways as of the Reactome database (31). Randomly sampling such sets, we observed that proteins that occur in many different pathways are preferable targets of any viral family (**Fig. 2C**). Recently, the focus of modern network research has shifted to the determination of nodes that allow the control of a network (39, 40). Notably, such genes were enriched with essential genes, disease genes as well as appeared in regulatory interactions (41, 42). Furthermore, they also played a role as targeted and required genes of viral infections (43). Determining proteins that always appeared in a control configuration termed critical nodes, we observed that targets of different families were enriched in such sets, while we found the opposite when we considered redundant proteins that never appear in such configurations (**Fig. 2D**).

Such topological considerations are not limited to single nodes as targets, required and essential genes appear to be weaved together in a dense network of connections (44, 45). While understudied in the context of host-virus interactions, recent advances suggest the existence of disease modules in molecular interaction networks, defined as a group of proteins in the same network vicinity . Several studies identified modules un- derlying diseases such as obesity and type 2 diabetes, asthma and inflammatory and malignant diseases (46-48) demonstrated that the identified disease modules partially overlap, sharing mo- lecular mechanisms as well as proteins. It was further shown that the degree of overlap correlates with biological similarity, disease symptoms and increased evidence of comorbidity (49). Therefore, the analysis of disease modules, combined with diverse data types such as genomics, genome-wide association studies, gene expression, gene- disease association, clinically relevant information may indicate a way to find affected molecular mechanisms, biomarkers and new drug targets for the analysis and treatment of viral infections. However, translation of the concept of disease modules on the host-pathogen relationship still remains problematic due to a lack of genome-wide interaction networks comprising both host and pathogen. Yet, recent advances Currently such models are restricted to gene subsets or describe gene islands (45).

Apart from topological considerations, we expect that different types of viruses and their families will utilize targets of similar, yet different functional classes. In **Fig. 3**, we utilized COGs classes of protein functions (50, 51). We determined the frequencies of such classes in sets of targets of different viral families and compared them to a profile of frequencies of functions of all human targets. The heatmap in **Fig. 3** suggests that papillomaviridae tend to strongly target various metabolic functions, while retroviridae and orthomyxoviridae show the opposite behavior. In turn, retroviridae mostly intervene in transcriptional functions, cell signaling and cell cycle control. Such differences have already been indicated previously, suggesting that different viral families use different strategies to invade a human host cell (45, 52).

Navratil et al. (2011) described a human infectome network (HIN) that linked 416 viral proteins to 1,148 human proteins through 2,099 manually curated virus-host PPIs (53). In fact, 32% of these cellular proteins are targeted by more than one virus protein. A similar fraction, 28% of these cellular targets interact with proteins from more than one virus.

Clearly, virus protein attack a relatively small number of human proteins that are relevant for their replication. More specifically, these human targets appear to be highly connected: the mean degree of these targets was 38 vs. 10 in non-targeted proteins (53). Even among highly connected proteins (k>5) in the human interactome, the degree of virus targets was twice as large as those of non-targeted proteins. Note that only 50% of all human proteins were known to interact with other proteins in the 2011 human proteome, and 50% of these interacting proteins were interacting with only one other protein. [do we have an update for this?]

**Essential host proteins and their interactions**

More than 1500 host proteins are known that are essential for virus infection and replication (53). However, only 11% of these are known to interact with virus proteins. However, about 42% of essential host factors (EHFs) were found as interactors of virus targets.

**The virus interactome-diseasome connection**

It has been long known that some viruses are involved in diseases not typically associated with infection. For instance, up to 20% of cancers may be caused by viruses such as papilloma or certain herpesviruses (54). Navratil et al used their list of virus targets and compared it to a list of 1729 human genetic disease-related proteins (derived from OMIM). It turns out that 13% of human virus targets are also associated with at least one human disease (53). That is, a human protein interacting with a virus protein is twice as likely to be involved in a disease than a non-target. Most of the diseases found in this study were related to cancer or neurodegenerative diseases. Surprisingly, type 1 diabetes was also associated with virus infection, as were autoimmune diseases in general. The latter may not be surprising, given that many virus infections elicit a strong immune reaction.

An independent study came to a similar conclusion: Gulbahce et al. analyzed the connection between Epstein-Barr-Virus (EBV), human papilloma virus (HPV) and disease (55). However, these authors not just used PPIs but also metabolic networks and regulatory interactions. Using U.S. Medicare patient medical history data derived from 13 million patients, Gulbahce et al. found that many diseases are often associated with viral infection, including EBV or HPV. For instance, HPV patients have 15.7 and 2.7 times increased chance of developing retina and bladder cancer but also a higher risk of Fanconi anemia (55).

**Virus interactions with the host metabolome**

There is increasing evidence that viruses not just highjack the host replication machinery, but also the host metabolic machinery. For instance, Adenovirus 5 proteins E4ORF1 and E4ORF6 co-immunoprecipitate with MYC in the nucleus, probably by directly interacting with MYC. While MYC has diverse effects on numerous target genes that it regulates, E4ORF1 induces MYC to activate a subset of glycolytic targets (viruses with a deletion of the E4 protein are defective for inducing glycolysis and decreased respiration). Thai et al. (56) conclusively demonstrated that adenovirus induced glycolysis generates metabolites for increased nucleotide biosynthesis in infected cells.

It is not surprising that viruses manipulate host metabolism to generate more nucleotides and other compounds that they need for replication (57). However, in most cases it remains unclear whether virus proteins directly or indirectly reprogram metabolism.

**Conclusions and outlook**

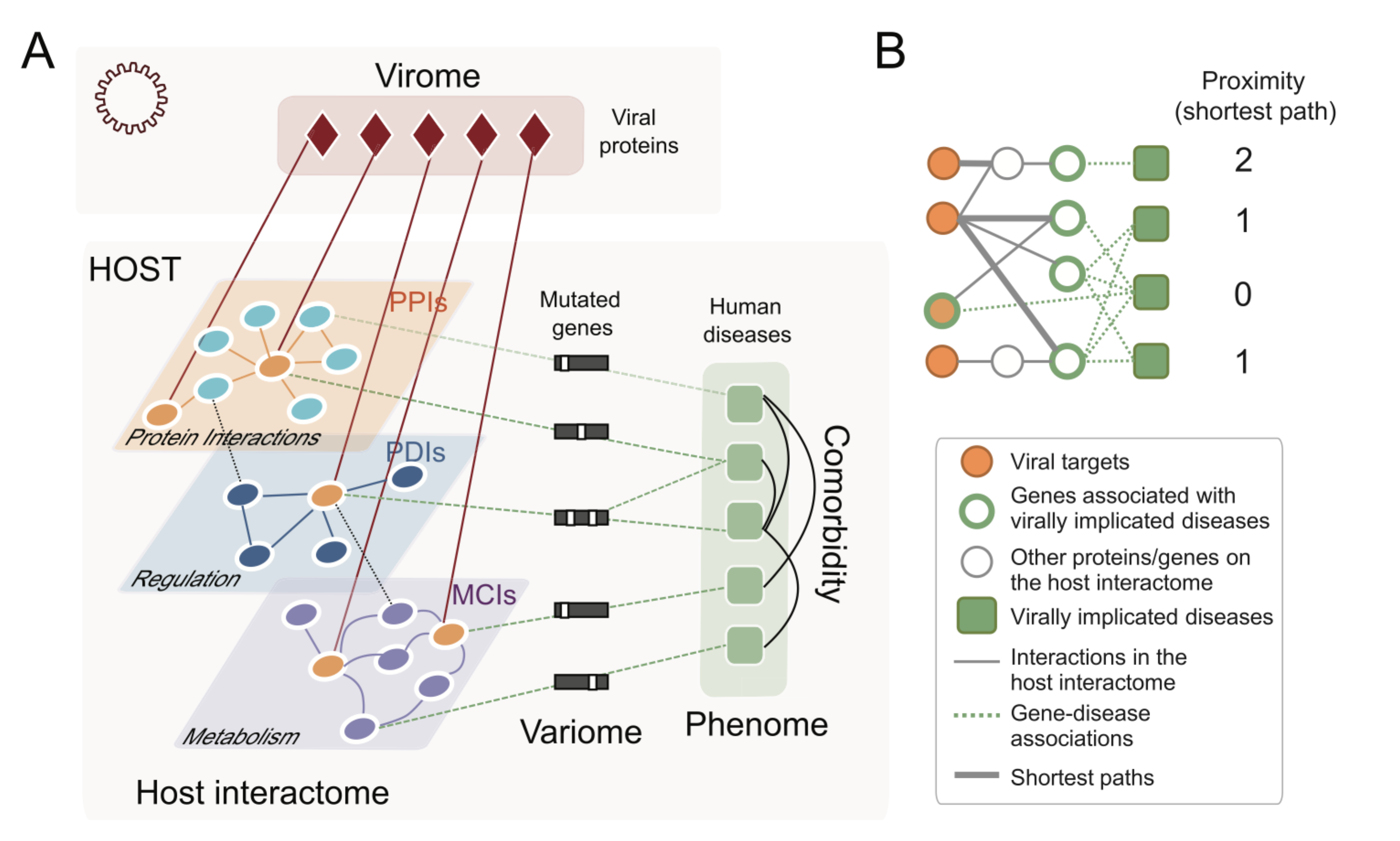
Protein-protein interactions are at the core of any virus infection, hence a detailed understanding of such interactions is critical for understanding viral diseases but also critical for the development of new drugs. For a small group of viruses we have so much interaction data that it has become difficult to evaluate their physiological significance, given that most databases do not provide such a critical judgement. For many viruses only few interactions are known and many more are expected to be detected.

Viruses are special in that they evolve much quicker than their hosts, especially in RNA viruses, hence the first line of defense in animals is their immune system which can cope in a tug-of-war with viruses only because they have both an efficient innate immune response and a highly flexible antibody production system (58).

With a rapidly growing number of human viruses, especially resulting from microbiome studies, we will identify many more viruses in humans. In the course of these studies, we will also find many more commensal viruses which do interact with their human host but may actually be beneficial, and potentially even help us to fight other pathogens and parasites.

**Figures**

**Figure 1: Overlaps of target sets of different virus families.** Assessing virus families by the number of available interactions between viral and human host proteins, we considered 1,988 targets of herpesviridae, 1,624 of orthomyxoviridae, 1,740 of papillomaviridae, 1,359 of retroviridae and a pool of 3,301 targets of other virus families. Considering the number of shared human targets between families, the Venn diagram suggests that targets of different families considerably overlap.

****

**Fig. 2 ?. From Gulbahce et al. 2012**

****

**Figure 2: Topological features of targets of herpesviridae, orthomyxoviridae, papillomaviridae, retroviridae and other families. (A)** Utilizing sets of proteins that are targeted by viruses of different families, we determined the enrichment of such target proteins in bins of proteins that have a certain number of interactions in a human protein-protein interaction network. Randomly sampling sets of viral targets, we observed that proteins of increasing degree are preferably targeted by viuses. **(B)** We defined a set of bottleneck proteins in a human protein-protein interaction network as the top 20% of proteins with highest betweeness centrality. Utilizing the set of proteins that were targeted by other virus families, set we observed 687 targeted bottleneck proteins. Randomized sampling of target sets confimed the statistical significance of the observed value, suggesting that bottleneck proteins are prime viral targets (P < 10-4). In the inset, we corrobaorate this observation, considering targets in remaining virus families. In **(C)**, we determined the occurrence of proteins in different pathways. Randomly sampling viral targets, we observed that targets of different virus families tend to appear in an increasing number of different pathways. In **(D)**, we determined the number of targeted critical, intermittent and redundant proteins. Randomizing such targets, we observed that critical proteins significantly accrued viral targets while we found the opposite for redundant proteins considering all virus families.



**Figure 3: Functions of targets of different virus families.** We determined the frequency of proteins that were targeted by different virus families. Such frequencies were compared to the corresponding frequencies of a set of targets of all families, determining a foldchange.

**Table 1: Overview of host-pathogen databases that provide human host-virus protein interactions.**

|  |  |  |  |
| --- | --- | --- | --- |
| **database** | **pathogens** | **webpage** | **References** |
| HCVPro | only HCV | http://www.cbrc.kaust.edu.sa/hcvpro/ | (17) |
| HIV-1 @NCBI | only HIV | https://www.ncbi.nlm.nih.gov/genome/viruses/retroviruses/hiv-1/interactions/ | (18) |
| PHIDIAS |  | http://www.phidias.us | (21) |
| PHISTO |  | http://www.phisto.org | (23) |
| HPIDB |  | http://www.agbase.msstate.edu/hpi/main.html | (22) |
| VirHostNet |  | http://virhostnet.prabi.fr | (19) |
| VirusMentha |  | http://virusmentha.uniroma2.it | (20) |

**Table 2**. **The 20 best-studied human viruses (by number of genomes sequenced)**. Sequence numbers as of July, 2016. Clustered sequenced were clustered at ≥98% sequence identity). C/D = un-/ clustered. Genome data from Genbank.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Type | Family name | Sequences (unclust.) | Disease (examples) | C/D | Total complete genomes | Complete genomes (clustered ) |
| RNA | reoviridae | 65870 | Rare diarrhea | 5.78 | 31945 | 5803 |
| RNA | flaviviridae | 225112 | Zika | 3.20 | 7837 | 2019 |
| DNA | hepadnaviridae | 78558 | hepatitis | 8.16 | 7248 | 1946 |
| plant | geminiviridae | 13158 | --- | 3.02 | 6421 | 2316 |
| RNA | picornaviridae | 85636 | Cold etc | 3.60 | 3447 | 1500 |
| RNA | retroviridae | 716088 | AIDS etc | 2.21 | 2890 | 2103 |
| anim | circoviridae | 7838 | --- | 6.92 | 2706 | 542 |
| RNA | phenuiviridae | 4139 | Rift Valley fever | 4.86 | 1678 | 384 |
| RNA | coronaviridae | 19164 | SARS | 5.61 | 1549 | 320 |
| RNA | potyviridae | 16115 |  | 3.63 | 1536 | 843 |
| DNA | papillomaviridae | 17847 | Warts, cancer | 8.14 | 1364 | 359 |
| DNA | polyomaviridae | 8604 | Rare cancers | 10.60 | 1277 | 164 |
| RNA | filoviridae | 2165 | Ebola | 23.53 | 1259 | 37 |
| RNA | togaviridae | 8924 | rubella | 16.25 | 1239 | 137 |
| RNA | pneumoviridae | 22578 | Cold-like | 13.08 | 1231 | 61 |
| plant | nanoviridae | 3110 | --- | 4.92 | 1183 | 282 |
| RNA | caliciviridae | 32405 | gastroenteritis | 5.30 | 1072 | 292 |
| RNA | paramyxoviridae | 29726 | measles | 10.06 | 1008 | 327 |
| RNA | bromoviridae | 4677 | (plants) | 4.08 | 764 | 384 |
| RNA | arenaviridae | 2639 | (animals) | 2.29 | 758 | 469 |

Statistics of viruses known to infect humans, genomes sequenced, genetic diversity?

**Table 3**. human disease caused by these viruses, case numbers, mortality, economic damage etc. Unless otherwise indicated, figures are yearly.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Virus (class)** | **Infections** | **Morbidity** | **Mortality** | **Cost** | **Refs** |
| HSV-1/2 | 3.7 B / ~700M | 3M/yr (US) | low | $540M (US) | (59-63) |
| HIV-1/2 | 36M (world) | 2.1M/yr ww | 25M total1 | $13.7B (US) | (64) |
| Influenza | >30M (US)3 | 30M (US)3 | 50M 19182 | $10-90B | (65-67) |
| Measles | >20 M (ww) | 250k ww | 140-500k4 | $3-7B (US) | (68) |
| Hepatitis C | 60-120M (ww) | 4M | 500k ww | $10B 5 | (69, 70) |
| Hepatitis B | 248 M ww/yr, ~2.5 B ww total | 350M ww total | 600k ww | $1B (US) | (71-73) |
| Zika | 740k S Amer | >2,6k 6 | low |  | (74) |
| MERS-CoV | 2067 |  | 720 total | $15-20B | (75) |
| SARS-CoV | 8098 |  | 774 total | $40B ww | (76, 77) |
| Common cold (rhinovirus) | 1B Cold/year (US) | 10-40% of common colds | low | $20B (US) | (78) |
| Norovirus (gastroenteritis) | 19-21M (US); 685M (ww) | 699M ww | 570-800 (US); 200K children ww  219K ww | $4.2B (indirect); $60.3B total ww | (79, 80) |

1 globally, since 1981. 2 Spanish flu of 1918. 3 30 million outpatient visits. 4 The death rate is decreasing, from 535,000 deaths in 2000 to 139,300 deaths in 2010. 5 $10·7 billion in direct medical expenditures in the USA for HCV-related disease from 2010 to 2019. 6 cases of microcephaly. K,M,B = thousand, million, billion, WW = worldwide, SA = South America.

**Table 4: Number of host-virus interactions of virus families.**

|  |  |  |
| --- | --- | --- |
| **viral family** | **# HPIs** | **viruses** |
| orthomyxoviridae | 5,495 |  |
| herpesviridae | 5,423 |  |
| papillomaviridae | 3,927 |  |
| retroviridae | 2,285 |  |
| paramyxoviridae | 873 |  |
| flaviviridae | 575 |  |
| reduviidae | 443 |  |
| poxviridae | 415 |  |
| polyomaviridae | 322 |  |
| parvoviridae | 292 |  |
| adenoviridae | 281 |  |
| cupressaceae | 247 |  |
| chromatiaceae | 235 |  |
| brassicaceae | 233 |  |
| filoviridae | 172 |  |
| bunyaviridae | 159 |  |
| iguanidae | 156 |  |
| togaviridae | 126 |  |
| rubiaceae | 108 |  |
| hepadnaviridae | 99 |  |
| asteromonadaceae | 88 |  |
| peribunyaviridae | 85 |  |
| nostocaceae | 82 |  |
| chironomidae | 76 |  |
| phenuiviridae | 76 |  |
| chloroflexaceae | 74 |  |
| arteriviridae | 67 |  |
| coronaviridae | 66 |  |
| reoviridae | 63 |  |
| fulgoridae | 61 |  |
| actinidiaceae | 60 |  |
| chattonellaceae | 51 |  |
| bradyrhizobiaceae | 33 |  |
| chlorobiaceae | 29 |  |
| other | 183 |  |

**References**

1. Uetz P*, et al.* (2006) Herpesviral protein networks and their interaction with the human proteome. *Science* 311(5758):239-242.

2. Zhang L*, et al.* (2009) Analysis of vaccinia virus-host protein-protein interactions: validations of yeast two-hybrid screenings. *J Proteome Res* 8(9):4311-4318.

3. Hauser R*, et al.* (2012) Bacteriophage protein-protein interactions. *Adv Virus Res* 83:219-298.

4. Blasche S, Wuchty S, Rajagopala SV, & Uetz P (2013) The protein interaction network of bacteriophage lambda with its host, Escherichia coli. *J Virol* 87(23):12745-12755.

5. de Chassey B*, et al.* (2008) Hepatitis C virus infection protein network. *Mol Syst Biol* 4:230.

6. Tripathi LP*, et al.* (2010) Network based analysis of hepatitis C virus core and NS4B protein interactions. *Mol Biosyst* 6(12):2539-2553.

7. Dolan PT*, et al.* (2013) Identification and comparative analysis of hepatitis C virus-host cell protein interactions. *Mol Biosyst* 9(12):3199-3209.

8. Ngo HT, Pham LV, Kim JW, Lim YS, & Hwang SB (2013) Modulation of mitogen-activated protein kinase-activated protein kinase 3 by hepatitis C virus core protein. *J Virol* 87(10):5718-5731.

9. Gautier VW*, et al.* (2009) In vitro nuclear interactome of the HIV-1 Tat protein. *Retrovirology* 6:47.

10. Jager S*, et al.* (2011) Global landscape of HIV-human protein complexes. *Nature* 481(7381):365-370.

11. Shapira SD*, et al.* (2009) A physical and regulatory map of host-influenza interactions reveals pathways in H1N1 infection. *Cell* 139(7):1255-1267.

12. Fossum E*, et al.* (2009) Evolutionarily conserved herpesviral protein interaction networks. *PLoS Pathog* 5(9):e1000570.

13. Calderwood MA*, et al.* (2007) Epstein-Barr virus and virus human protein interaction maps. *Proc Natl Acad Sci U S A* 104(18):7606-7611.

14. Forsman A, Ruetschi U, Ekholm J, & Rymo L (2008) Identification of intracellular proteins associated with the EBV-encoded nuclear antigen 5 using an efficient TAP procedure and FT-ICR mass spectrometry. *J Proteome Res* 7(6):2309-2319.

15. Khadka S*, et al.* (2011) A physical interaction network of dengue virus and human proteins. *Mol Cell Proteomics* 10(12):M111 012187.

16. Pichlmair A*, et al.* (2012) Viral immune modulators perturb the human molecular network by common and unique strategies. *Nature* 487(7408):486-490.

17. Kwofie SK, Schaefer U, Sundararajan VS, Bajic VB, & Christoffels A (2011) HCVpro: hepatitis C virus protein interaction database. *Infect Genet Evol* 11(8):1971-1977.

18. Ako-Adjei D*, et al.* (2015) HIV-1, human interaction database: current status and new features. *Nucleic Acids Res* 43(Database issue):D566-570.

19. Guirimand T, Delmotte S, & Navratil V (2015) VirHostNet 2.0: surfing on the web of virus/host molecular interactions data. *Nucleic Acids Res* 43(Database issue):D583-587.

20. Calderone A, Licata L, & Cesareni G (2015) VirusMentha: a new resource for virus-host protein interactions. *Nucleic Acids Res* 43(Database issue):D588-592.

21. Xiang Z, Tian Y, & He Y (2007) PHIDIAS: a pathogen-host interaction data integration and analysis system. *Genome Biol* 8(7):R150.

22. Kumar R & Nanduri B (2010) HPIDB--a unified resource for host-pathogen interactions. *BMC Bioinformatics* 11 Suppl 6:S16.

23. Durmus Tekir S*, et al.* (2013) PHISTO: pathogen-host interaction search tool. *Bioinformatics* 29(10):1357-1358.

24. Aranda B*, et al.* (2011) PSICQUIC and PSISCORE: accessing and scoring molecular interactions. *Nat Methods* 8(7):528-529.

25. Alfarano C*, et al.* (2005) The Biomolecular Interaction Network Database and related tools 2005 update. *Nucleic Acids Res* 33(Database issue):D418-424.

26. Chatr-Aryamontri A*, et al.* (2013) The BioGRID interaction database: 2013 update. *Nucleic Acids Res* 41(Database issue):D816-823.

27. Salwinski L*, et al.* (2004) The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res* 32(Database issue):D449-451.

28. Keshava Prasad TS*, et al.* (2009) Human Protein Reference Database--2009 update. *Nucleic Acids Res* 37(Database issue):D767-772.

29. Orchard S*, et al.* (2014) The MIntAct project--IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res* 42(Database issue):D358-363.

30. Licata L*, et al.* (2012) MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res* 40(Database issue):D857-861.

31. Croft D*, et al.* (2014) The Reactome pathway knowledgebase. *Nucleic Acids Res* 42(Database issue):D472-477.

32. Das J, Mohammed J, & Yu H (2012) Genome-scale analysis of interaction dynamics reveals organization of biological networks. *Bioinformatics* 28(14):1873-1878.

33. Durmus Tekir SD & Uelgen K (2013) Systems biology of pathogen-host interaction: Networks of protein-protein interaction within pathogens and pathogen-human interactions in the post-genomic era. *Biotechn. J.* 8(1):85-96.

34. Arnold R, Boonen K, Sun MG, & Kim PM (2012) Computational analysis of interactomes: current and future perspectives for bioinformatics approaches to model the host-pathogen interaction space. *Methods* 57(4):508-518.

35. Korth MJ, Tchitchek N, Benecke AG, & Katze MG (2013) Systems approaches to influenza-virus host interactions and the pathogenesis of highly virulent and pandemic viruses. *Semin Immunol* 25(3):228-239.

36. Nourani E, Khunjush F, & Durmus S (2015) Computational approaches for prediction of pathogen-host protein-protein interactions. *Front Microbiol* 6:94.

37. Dyer MD, Murali TM, & Sobral BW (2011) Supervised learning and prediction of physical interactions between human and HIV proteins. *Infect Genet Evol* 11(5):917-923.

38. Dyer MD, Murali TM, & Sobral BW (2008) The landscape of human proteins interacting with viruses and other pathogens. *PLoS Pathog* 4(2):e32.

39. Ishitsuka M, Akutsu T, & Nacher JC (2016) Critical controllability in proteome-wide protein interaction network integrating transcriptome. *Scientific reports* 6:23541.

40. Nacher JC & Akutsu T (2014) Analysis of critical and redundant nodes in controlling directed and undirected complex networks using dominating sets. *J. Compl. Networks* 2(4):394-412.

41. Wuchty S (2014) Controllability in protein interaction networks. *Proc Natl Acad Sci U S A* 111(19):7156-7160.

42. Khuri S & Wuchty S (2015) Essentiality and centrality in protein interaction networks revisited. *BMC bioinformatics* 16:109.

43. Wuchty S, Boltz T, & Kucuk-McGinty H (2017) Links between critical proteins drive the controllability of protein interaction networks. *Proteomics*.

44. Mariano R, Khuri S, Uetz P, & Wuchty S (2016) Local Action with Global Impact: Highly Similar Infection Patterns of Human Viruses and Bacteriophages. *mSystems* 1(2).

45. Dix A, Vlaic S, Guthke R, & Linde J (2016) Use of systems biology to decipher host-pathogen interaction networks and predict biomarkers. *Clin Microbiol Infect* 22(7):600-606.

46. Sharma A*, et al.* (2015) A disease module in the interactome explains disease heterogeneity, drug response and captures novel pathways and genes in asthma. *Hum Mol Genet* 24(11):3005-3020.

47. Barrenas F*, et al.* (2012) Highly interconnected genes in disease-specific networks are enriched for disease-associated polymorphisms. *Genome Biol* 13(6):R46.

48. Gustafsson M*, et al.* (2014) Integrated genomic and prospective clinical studies show the importance of modular pleiotropy for disease susceptibility, diagnosis and treatment. *Genome Med* 6(2):17.

49. Menche J*, et al.* (2015) Disease networks. Uncovering disease-disease relationships through the incomplete interactome. *Science* 347(6224):1257601.

50. Tatusov RL*, et al.* (2003) The COG database: an updated version includes eukaryotes. *BMC bioinformatics* 4:41.

51. Franceschini A*, et al.* (2013) STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic acids research* 41(Database issue):D808-815.

52. Brito AF & Pinney JW (2017) Protein-Protein Interactions in Virus-Host Systems. *Front Microbiol* 8:1557.

53. Navratil V, de Chassey B, Combe CR, & Lotteau V (2011) When the human viral infectome and diseasome networks collide: towards a systems biology platform for the aetiology of human diseases. *BMC Syst Biol* 5:13.

54. Morales-Sanchez A & Fuentes-Panana EM (2014) Human viruses and cancer. *Viruses* 6(10):4047-4079.

55. Gulbahce N*, et al.* (2012) Viral perturbations of host networks reflect disease etiology. *PLoS Comput Biol* 8(6):e1002531.

56. Thai M*, et al.* (2014) Adenovirus E4ORF1-induced MYC activation promotes host cell anabolic glucose metabolism and virus replication. *Cell Metab* 19(4):694-701.

57. Miyake-Stoner SJ & O'Shea CC (2014) Metabolism goes viral. *Cell Metab* 19(4):549-550.

58. Christiaansen A, Varga SM, & Spencer JV (2015) Viral manipulation of the host immune response. *Curr Opin Immunol* 36:54-60.

59. Boppana SB & Fowler KB (2007) Persistence in the population: epidemiology and transmisson. *Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis*, eds Arvin A, Campadelli-Fiume G, Mocarski E, Moore PS, Roizman B, Whitley R, & Yamanishi KCambridge).

60. Burrel S*, et al.* (2017) Ancient Recombination Events between Human Herpes Simplex Viruses. *Mol Biol Evol* 34(7):1713-1721.

61. Johnston C, Gottlieb SL, & Wald A (2016) Status of vaccine research and development of vaccines for herpes simplex virus. *Vaccine* 34(26):2948-2952.

62. Looker KJ*, et al.* (2015) Global and Regional Estimates of Prevalent and Incident Herpes Simplex Virus Type 1 Infections in 2012. *PLoS One* 10(10):e0140765.

63. Szucs TD, Berger K, Fisman DN, & Harbarth S (2001) The estimated economic burden of genital herpes in the United States. An analysis using two costing approaches. *BMC Infect Dis* 1:5.

64. Menzies NA*, et al.* (2011) The cost of providing comprehensive HIV treatment in PEPFAR-supported programs. *AIDS* 25(14):1753-1760.

65. Johnson NP & Mueller J (2002) Updating the accounts: global mortality of the 1918-1920 "Spanish" influenza pandemic. *Bull Hist Med* 76(1):105-115.

66. Merson MH, O'Malley J, Serwadda D, & Apisuk C (2008) The history and challenge of HIV prevention. *Lancet* 372(9637):475-488.

67. Molinari NA*, et al.* (2007) The annual impact of seasonal influenza in the US: measuring disease burden and costs. *Vaccine* 25(27):5086-5096.

68. Patel MK*, et al.* (2016) Progress Toward Regional Measles Elimination - Worldwide, 2000-2015. *MMWR Morb Mortal Wkly Rep* 65(44):1228-1233.

69. Wong JB, McQuillan GM, McHutchison JG, & Poynard T (2000) Estimating future hepatitis C morbidity, mortality, and costs in the United States. *Am J Public Health* 90(10):1562-1569.

70. Wang H & Collaborators GBoD (2016) Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* 388(10053):1459-1544.

71. Maynard JE (1990) Hepatitis B: global importance and need for control. *Vaccine* 8 Suppl:S18-20; discussion S21-13.

72. Ott JJ, Stevens GA, Groeger J, & Wiersma ST (2012) Global epidemiology of hepatitis B virus infection: new estimates of age-specific HBsAg seroprevalence and endemicity. *Vaccine* 30(12):2212-2219.

73. Keshavarz K*, et al.* (2015) Economic burden of hepatitis B virus-related diseases: evidence from iran. *Hepat Mon* 15(4):e25854.

74. Fischer M (2016) Zika virus epidemiology update.

75. Tang XC*, et al.* (2014) Identification of human neutralizing antibodies against MERS-CoV and their role in virus adaptive evolution. *Proc Natl Acad Sci U S A* 111(19):E2018-2026.

76. Oberholtzer K (2004) *Learning from SARS: Preparing for the Next Disease Outbreak--Workshop Summary.* (National Academies Press).

77. Lee J-W & McKibbin WJ (2004) Estimating the global economic costs of SARS. *Learning from SARS: preparing for the next disease outbreak: workshop summary.,* Institute of Medicine (US) Forum on Microbial Threats., (National Academies Press, WAshinton, DC).

78. Simasek M & Blandino DA (2007) Treatment of the common cold. *Am Fam Physician* 75(4):515-520.

79. Bartsch SM, Lopman BA, Ozawa S, Hall AJ, & Lee BY (2016) Global Economic Burden of Norovirus Gastroenteritis. *PLoS One* 11(4):e0151219.

80. Robilotti E, Deresinski S, & Pinsky BA (2015) Norovirus. *Clin Microbiol Rev* 28(1):134-164.