The Landscape of Human Proteins Interacting with Viruses and Other Pathogens

Matthew D. Dyer[1,2], T. M. Murali[3*], Bruno W. Sobral[2*]

1 Genetics, Bioinformatics, and Computational Biology Program, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, United States of America, 2 Virginia Bioinformatics Institute, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, United States of America, 3 Department of Computer Science, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, United States of America

Infectious diseases result in millions of deaths each year. Mechanisms of infection have been studied in detail for many pathogens. However, many questions are relatively unexplored. What are the properties of human proteins that interact with pathogens? Do pathogens interact with certain functional classes of human proteins? Which infection mechanisms and pathways are commonly triggered by multiple pathogens? In this paper, to our knowledge, we provide the first study of the landscape of human proteins interacting with pathogens. We integrate human–pathogen protein–protein interactions (PPIs) for 190 pathogen strains from seven public databases. Nearly all of the 10,477 human-pathogen PPIs are for viral systems (98.3%), with the majority belonging to the human–HIV system (77.9%). We find that both viral and bacterial pathogens tend to interact with hubs (proteins with many interacting partners) and bottlenecks (proteins that are central to many paths in the network) in the human PPI network. We construct separate sets of human proteins interacting with bacterial pathogens, viral pathogens, and those interacting with multiple bacteria and with multiple viruses. Gene Ontology functions enriched in these sets reveal a number of processes, such as cell cycle regulation, nuclear transport, and immune response that participate in interactions with different pathogens. Our results provide the first global view of strategies used by pathogens to subvert human cellular processes and infect human cells. Supplementary data accompanying this paper is available at http://staff.vbi.vt.edu/dyermd/publications/dyer2008a.html.

Introduction

Infectious diseases result in millions of deaths each year. Millions of dollars are spent annually to better understand how pathogens infect their hosts and to identify potential targets for therapeutics. An important aspect of any host-pathogen system is the mechanism by which a pathogen is able to invade a host cell. Within these complex systems, protein-protein interactions (PPIs) between surface proteins form the foundation of communication between a host and a pathogen and play a vital role in initiating infection [1]. PPI-mediated mechanisms of infection have been studied in detail for many pathogens [2–7]. However, many questions are relatively unexplored. What are the properties of human proteins that interact with pathogens? Do pathogens interact with certain functional classes of human proteins? Which infection mechanisms and pathways are commonly triggered by multiple pathogens? A significant hurdle to such global cross-pathogen comparisons has been the shortage of large-scale datasets of interactions between host and pathogen proteins. High-throughput experimental screens have been primarily used to identify intraspecies PPIs [8–16]. However, recent efforts to include host-pathogen PPIs in public databases have made it easier to acquire the data needed to address these important questions.

In this paper, we integrate experimentally verified human-pathogen PPIs for 190 pathogen strains from seven public databases [17–23]. We partition the strains into 54 different pathogen groups, where each group is made up of taxonomically related strains. We analyze the intraspecies network of PPIs between the 1,233 unique human proteins spanned by the host-pathogen PPIs, and find that pathogens, both viral and bacterial, tend to interact with hubs (proteins with many interacting partners) and bottlenecks (proteins that are central to many paths in the network) in the human PPI network.

We pay special attention to two networks of PPIs between human proteins: the proteins that interact with at least two viral pathogen groups (see Figure 1) and the proteins that interact with at least two bacterial pathogen groups (see Figure 2, noting that the figure also contains human proteins targeted by only one bacterial pathogen group). We used the Cerebral plugin [24] for Cytoscape [25] to render these images. We compute the Gene Ontology (GO) [26] functions enriched in each of these two sets of human proteins. Such enriched functions highlight human pathways that may be involved in infection mechanisms that are common to multiple pathogens. Examples of such processes and components include cell cycle regulation, I-κB kinase/NF-κB cascade, and the nuclear membrane. These functions shed light on a number of features shared by different pathogens: interacting with human transcription factors and key proteins that control the cell cycle; transport of genetic material through the nuclear membrane (in the case of viruses) to subvert the host's transcriptional machinery;

* To whom correspondence should be addressed. E-mail: murali@cs.vt.edu (TMM), sobral@vbi.vt.edu (BWS)

Author Summary

Many pathogens, such as viruses and bacteria, cause disease in humans. Pathogen infections result in illness and death for millions of people each year. Pathogens communicate with human cells through physical interactions with various human proteins on the surface of the cell and within the interior of the cell. These interactions allow the pathogen to enter the host cell, manipulate important cellular processes, multiply, and invade other cells. In this paper, we compare interactions between human and pathogen proteins from 190 different pathogens to provide important insights into strategies used by pathogens to infect human cells. We show that both viral and bacterial proteins interact with human proteins that themselves interact with many human proteins or with human proteins that lie on many communication channels between other human proteins. Pathogens may have evolved to interact with these human proteins since they may control critical human cellular process. We also demonstrate that many viruses share common infection strategies, e.g., lengthening particular stages of the cell cycle, controlling programmed cell death, and interacting with the nuclear membrane to transfer viral genetic material into and out of the nucleus. Such studies may help us better understand the process of infection and identify better strategies to prevent or cure infection.

triggering an immune response via toll-like receptors; and activation of NF-κB signaling. We discuss in detail the importance of these and other enriched functions, as well as the proteins they annotate and the pathogens they interact with. Overall, these results provide the first global view of aspects of human cellular processes that are controlled by and respond to pathogens.

Our results should be interpreted with caution since no single pathogen may target all the proteins and PPIs we analyze. In addition, data for bacterial pathogens are scarce. However, we suggest that piecing together targeted human proteins across multiple pathogens has the potential to provide insights into common molecular mechanisms of infection and proliferation used by different pathogens.

Results/Discussion

We use the term "pathogen group" to refer to a set of pathogen strains that are closely related taxonomically, i.e., they all belong to the same genus, or, in the case of viruses, the same family. We partition the 190 strains into 54 pathogen groups: 35 viral, 17 bacterial, and two protozoan. Nearly all of the 10,477 human-pathogen PPIs we collect are for viral systems (98.3%), with the majority belonging to the human-HIV system (77.9%). These human-pathogen PPIs involve 1,233 unique human proteins, of which 1,109 are known to interact with at least one other human protein. Of these 1,233 human proteins, 221 interact with at least two pathogen groups (182 with more than one viral pathogen and 20 with more than one bacterial pathogen).

Pathogens Target Protein Hubs and Bottlenecks

Researchers have argued that the degree distribution of PPI networks is scale-free and follows the power law, i.e., the fraction of proteins in the network interacting with k other proteins is proportional to $k^{-\gamma}$, for some γ greater than zero, typically between two and three [27,28]. One feature of such

Figure 1. Human Proteins Interacting with Multiple Viral Pathogen Groups
The network of interactions between human proteins interacting with at least two viral pathogen groups. The size and color of a protein denote the number of pathogen groups that interact with it: light blue is two, dark blue is three, green is four, yellow is five, orange is six, and red is seven.
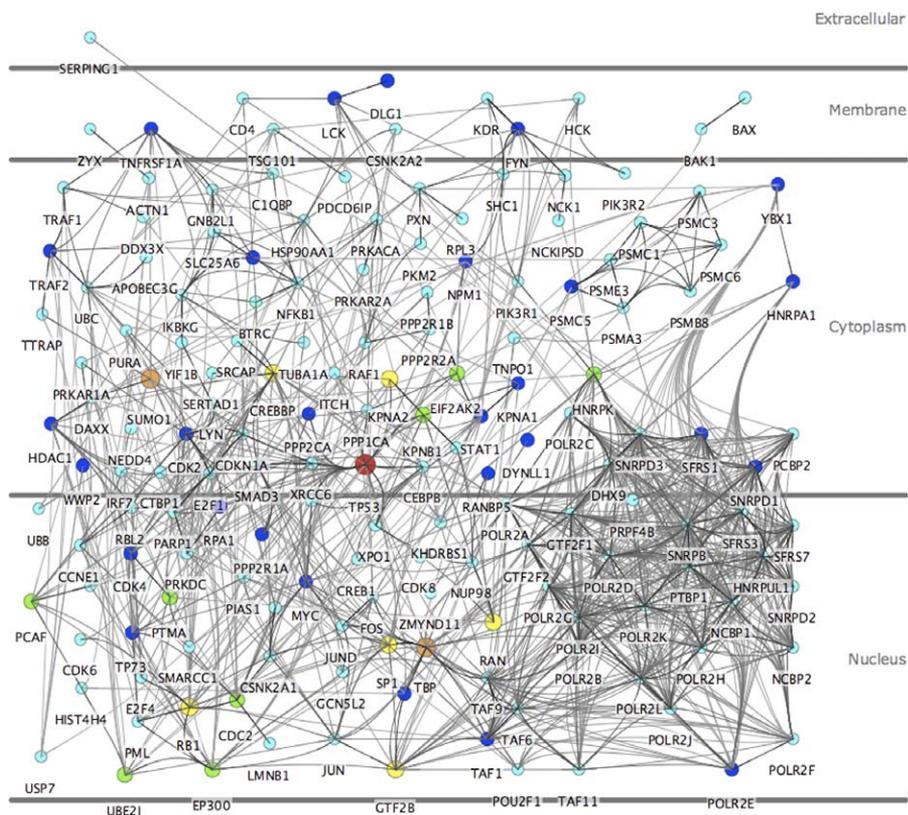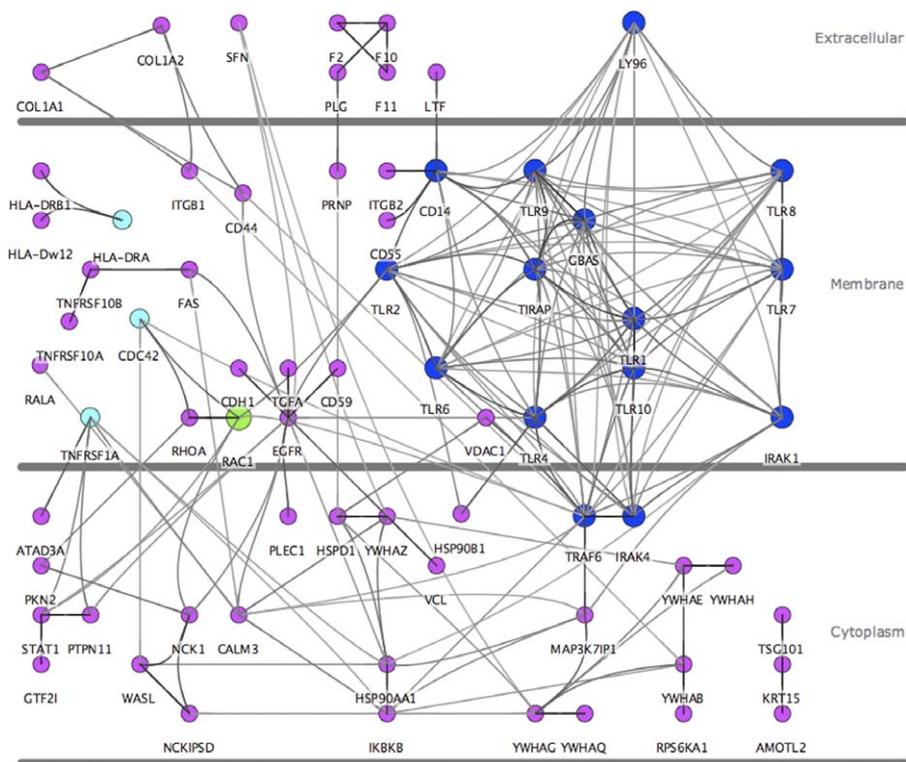doi:10.1371/journal.ppat.0040032.g001

Figure 2. Human Proteins Interacting with Bacterial Pathogen Groups

The network of interactions between human proteins interacting with at least one bacterial pathogen group. The size and color of a protein denote the number of pathogen groups that interact with it: purple is one, light blue is two, dark blue is three, and green is four.

doi:10.1371/journal.ppat.0040032.g002

networks is that they are robust in the face of attacks on random nodes. For instance, the removal of random subsets of nodes increases the diameter of the network only gradually [29,30]. In this context, the diameter is defined as the average length of the shortest paths between all pairs of proteins. However, the selective removal of even a small number of nodes of high degree can dramatically change the topology of the network [29,30].

There is considerable debate on the origins of the scale-free property and whether this property is an artifact of experimental biases and errors [31–33]. Notwithstanding this debate, we reasoned that pathogens may have evolved to interact with human proteins that are hubs (those involved in many interactions) or bottlenecks (those central to many pathways) [34] to disrupt key proteins in complexes and pathways. (See Methods for a precise definition of "bottleneck.") Our results support this hypothesis. Figure 3A displays the cumulative log-log plot of the degree distribution of four sets of proteins in the human PPI network: (i) all proteins, (ii) "Viral" set, the subset of proteins interacting with at least one viral pathogen group, (iii) "Bacterial" set, the subset of proteins interacting with at least one bacterial pathogen group, and (iv) "Multiviral" set, the subset of proteins interacting with at least two viral pathogen groups. We did not include the "Multibacterial" set of human proteins interacting with two or more bacterial pathogen groups in this analysis since there are only 20 such proteins. These plots show that across almost the entire range of degrees, proteins interacting with viral and bacterial pathogen groups tend to have higher degrees than human proteins

not interacting with pathogens. Further, proteins interacting with at least two viral pathogens have higher degrees than proteins interacting with one or more viral pathogens. The betweenness centrality results display the same trend (see Figure 3B). Across the entire range of values, proteins interacting with viral and bacterial pathogens have higher betweenness centrality. These results suggest that pathogens may have evolved to interact with human hub and bottleneck proteins, perhaps because these proteins control critical processes in the host cell.

We used Gene Set Enrichment Analysis (GSEA) [35] to test whether the gaps we observed in Figure 3 are statistically significant. GSEA is a method developed to assess the significance of the differential expression of a pre-defined gene set in two phenotypes of interest [35]. GSEA ranks all genes by a suitable measure of differential expression (e.g., the t-statistic) and uses a modified Kolmogorov-Smirnov test to assess if the genes in the given set have surprisingly high or low ranks. Since distributions of the t-statistics of differentially expressed genes have been observed to follow a power-law distribution [36], we reasoned that GSEA may be appropriate to test whether the human proteins interacting with pathogens have surprisingly high degree or betweenness centrality.

Our GSEA results support the conclusions we draw from Figure 3 that pathogens preferentially interact with human protein hubs and bottlenecks: for each of the three sets of proteins plotted in Figure 3, GSEA yields a p-value of at most 3×10^{-5} (degree) and 2.3×10^{-4} (centrality). To alleviate the concern that the observed patterns may be artifacts of
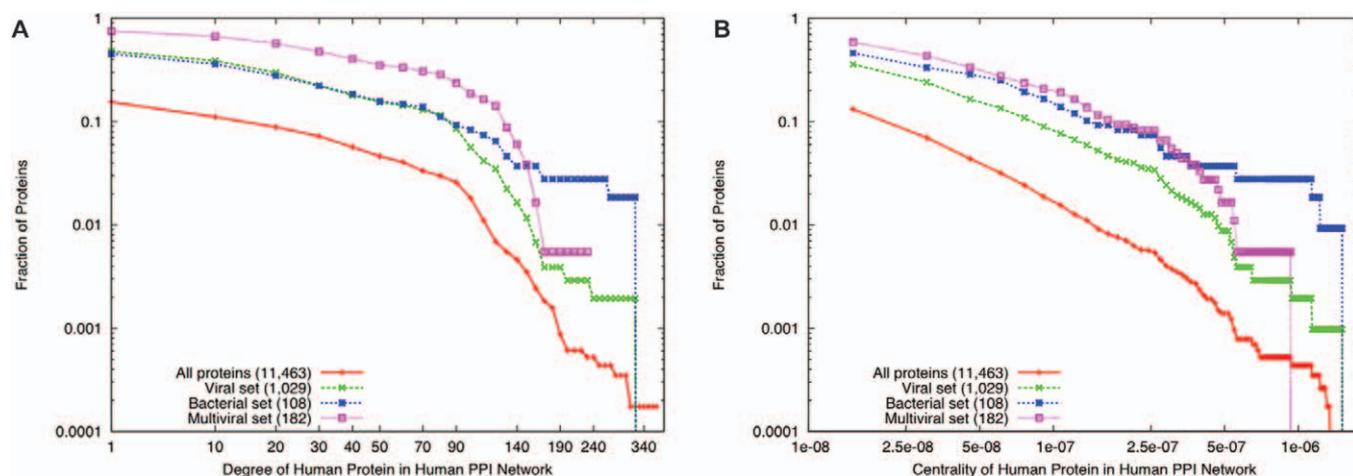
Figure 3. Degree and Centrality Distributions

Cumulative log-log distributions of (A) node degrees and (B) centralities for four subsets of nodes in the human PPI network: (i) red pluses are the set of all proteins in the network; (ii) green squares correspond to the viral set; (iii) blue crosses are for the bacterial set, and (iv) magenta squares are for the multiviral set. Numbers in parentheses represent the number of proteins in each set. The fraction of proteins at a particular value of degree or centrality is the number of proteins having that value or greater divided by the number of proteins in the set.

doi:10.1371/journal.ppat.0040032.g003

experimental biases or errors in the human PPI network, we repeated each of the analyses using two subsets of the human PPI network: a network composed of 13,324 PPIs detected only by high-throughput studies [14,15,37] and a network with 59,396 PPIs constructed using only manually curated interactions [20,23]. The top half of Table 1 summarizes these results. For all three networks, the viral set, the bacterial set, and the multiviral set are significant at the 0.05 level for both degree and centrality, with the exception of the multiviral set in the high-throughput network. Since 77.9% of the human-pathogen PPIs are for the human-HIV system, we repeated these analyses for each network after removing all human-HIV PPIs and obtained similar results (see the bottom half of Table 1). In Text S1, we discuss three analyses that show that

the consistency in the GSEA results for degree and for centrality are unlikely to result from any correlation that may exist between a protein's degree and its centrality (Figure S1 and Table S1 accompany the discussion in Text S1). We note that Tables S2 and S3 of the supplementary data contain detailed information on the GSEA results for the groups in Figure 3 and for individual pathogen groups.

Functions Enriched in Proteins Interacting with Pathogens

We computed over-represented GO terms in 58 sets of human proteins: the bacterial set, the viral set, the multi-bacterial set, the multiviral set, and the 54 sets of human proteins interacting with each of the 54 pathogen groups. Overall, we found 404 unique GO terms enriched in these

Table 1. GSEA Results

	Network	Protein Set	Number Proteins in Group	Degree		Centrality	
				ES	p-Value	ES	p-Value
All Human-Pathogen PPIs	W	Virus	1,029	0.79	$< 1 \times 10^{-6}$	0.83	$< 1 \times 10^{-6}$
		Multivirus	182	0.84	$< 1 \times 10^{-6}$	0.86	1.2×10^{-4}
		Bacteria	108	0.76	3×10^{-5}	0.89	2.3×10^{-4}
	HT	Virus	466	0.68	$< 1 \times 10^{-6}$	0.82	1.5×10^{-3}
		Multivirus	98	0.65	0.03	0.82	0.1
		Bacteria	43	0.79	2×10^{-3}	0.89	0.02
	MC	Virus	958	0.78	$< 1 \times 10^{-6}$	0.80	$< 1 \times 10^{-6}$
		Multivirus	174	0.83	$< 1 \times 10^{-6}$	0.82	1.9×10^{-5}
		Bacteria	100	0.73	3.4×10^{-4}	0.85	9.2×10^{-4}
Without Human-HIV PPIs	W	Virus	499	0.80	$< 1 \times 10^{-6}$	0.85	$< 1 \times 10^{-6}$
		Multivirus	81	0.83	$< 1 \times 10^{-6}$	0.88	3.6×10^{-3}
	HT	Virus	267	0.70	1×10^{-6}	0.84	6.12×10^{-4}
		Multivirus	46	0.72	0.02	0.86	0.07
	MC	Virus	958	0.78	$< 1 \times 10^{-6}$	0.80	$< 1 \times 10^{-6}$
		Multivirus	174	0.83	$< 1 \times 10^{-6}$	0.85	1.3×10^{-3}

Summary of GSEA results with and without human-HIV interactions for three networks: Whole human PPI network (W), the human PPI network yielded by High-Throughput experiments (HT), and the human PPI network consisting only of Manually Curated PPIs (MC). We report p-values only for the sets of human proteins in Figure 3. The "number of proteins in group" column displays the total number of human proteins in that group. The "ES" column displays the enrichment score calculated by GSEA.

doi:10.1371/journal.ppat.0040032.t001

sets. A complete list of enriched GO terms with images of the sub-networks spanned by the human proteins annotated with each term is available on the supplementary website.

We identified at least one enriched function in 21 pathogen groups. Analysis of these data identified 91 biclusters (see Methods for details), each containing between two and seven pathogen groups and between two and 40 enriched GO functions. We focus on two of the biclusters below. The biclusters demonstrate that our analysis can group different enriched functions together even if the effects of the interactions on the host cell or the participating host proteins are different.

Our first example is a bicluster spanning the three pathogen groups Adenovirus, HIV, and Papillomavirus and 23 GO functions. GO biological processes in the bicluster include "cell cycle process" and "regulation of cellular process." GO cellular components in the bicluster include "membrane-enclosed lumen" and "pore complex." The membrane-enclosed lumen is the space within a sealed membrane or between two sealed membranes. Proteins annotated with these functions include KPNA2, a karyopherin, the histone deacetylases HDAC1 and HDAC2, and a number of Transcription Factors (TFs). KPNA2 plays an important role in both the import and export of material through the nuclear membrane. Interactions with KPNA2 enable a virus to enter the nucleus and take over the host's transcriptional machinery [38–41]. HDACs play an important role in silencing gene expression by removing acetyl groups from histones, thus causing them to wrap more tightly around DNA and block the binding of TFs. The role played by pathogen-HDAC interactions varies among pathogen groups. In the case of Adenovirus, it has been suggested that the pathogen protein E1B interacts with HDAC1/SIN3 to produce an enzymatically active complex that may be capable of repressing the transcriptional activity of the human TP53 protein in order to block apoptosis [42]. In contrast, the E7 Papillomavirus protein binds to the HDAC complex to promote cell growth, eventually leading to cervical cancer [43].

The second example is a bicluster containing a virus (HIV) and three bacteria (Chlamydia, Neisseria, and *Escherichia coli*). This bicluster contains 11 GO functions including the biological processes "immune response," "response to stimulus," and "cytokine production." Although these four groups of pathogens interact with proteins belonging to the same pathways, the functions of the interactions are different. In the case of the bacteria, these functions annotate such proteins as toll-like receptors (TLRs) and interleukin receptor-associated kinases (IRAKs), which are special classes of host proteins responsible for recognizing foreign material and activating an immune response. There are no reported interactions with these proteins and HIV, although some researchers suggest that the single-stranded RNA of HIV-1 may encode many TLR7/TLR8 ligands [44]. In contrast to the bacteria in the bicluster, HIV uses host proteins involved in immune response such as CD4, CCR5, and CXCR4 to gain entrance to the cell. HIV attaches to the host protein CD4, a T cell glycoprotein, and subsequently to host chemokine receptors CCR5 and CXCR4. These binding events cause conformational changes to host proteins that allow the membrane of the virus to fuse to the host cell membrane [1].

The Network of Proteins Interacting with Multiple Pathogens

The biclustering analysis of the previous section suggests that specific sets of pathogen groups might trigger or target the same human pathways and processes. Encouraged by these data, we asked if there are infection pathways commonly targeted or triggered by at least two viral or bacterial pathogen groups. To answer this question, we constructed two networks of human proteins: one where every protein interacts with at least two viral pathogen groups and the other where every protein interacts with at least two bacterial pathogen groups. In each network, we included every PPI connecting two proteins in the network. Figures 1 and 2 display these networks. (Note that Figure 2 also contains human proteins that interact with only one bacterial pathogen group.) We computed the enriched GO functions in these two networks. We group and highlight some of the enriched functions and relevant sub-networks below. Throughout our discussion, we will refer to the localization of proteins in the four main regions of Figures 1 and 2: extracellular, the cell membrane, the cytoplasm, and the nucleus. For every GO function that we discuss, we mention its p-value and rank in the sorted list of all functions enriched in the corresponding network.

Human Proteins Targeted by Multiple Viral Pathogens

Our analysis highlights a number of important mechanisms that viral pathogens use to manipulate the human cell: (i) control the host cell cycle program to ensure the transcription of viral genetic material; (ii) utilize human TFs to promote the transcription of viral genetic material; (iii) target key human proteins that regulate critical cellular processes such as apoptosis; and (iv) subvert host machinery for transporting material across the nuclear membrane.

Control the host cell cycle program. Many viral pathogens are known to manipulate host cell cycle processes [45–47]. Our enrichment results reflect these findings. Our analysis identifies a sub-network of human proteins targeted by multiple viral pathogen groups enriched in the biological process "cell cycle" (p-value 6.2×10^{-6}, rank 21/89). Figure 4 displays this network. In this figure, we used GO annotations to clarify in which phase of the cell cycle each protein participates. The proteins in this figure are scattered through the cytoplasm and nucleus regions of Figure 1.

Two stages of the cell cycle are enriched in our analysis: "G_1 phase" (p-value 0.004, rank 52/89) and "Interphase" (p-value 0.01, rank 60/89). Images for these functions are available on the supplementary website. G_1 is the initial stage of the cell cycle. In this phase, a number of proteins needed for DNA replication are transcribed and translated. A direct link between pathogen interference and the G_1 phase has been established for HIV [48]. The HIV TAT protein elongates the G_1 phase in order to promote viral gene expression. Of the 13 human proteins in Figure 4 that participate in G_1, ten are known to interact with TAT. One of these interactions is with the human protein RB1, a retinoblastoma-associated protein and a known tumor suppressor, which can repress genes transcribed by the E2F family of transcription factors that are required for entering the S phase of the cell cycle [49]. RB1 interacts with five pathogens in total: Adenovirus, Herpesvirus, HIV, Papillomavirus, and Simian virus [50–54]. In the case of HIV, the
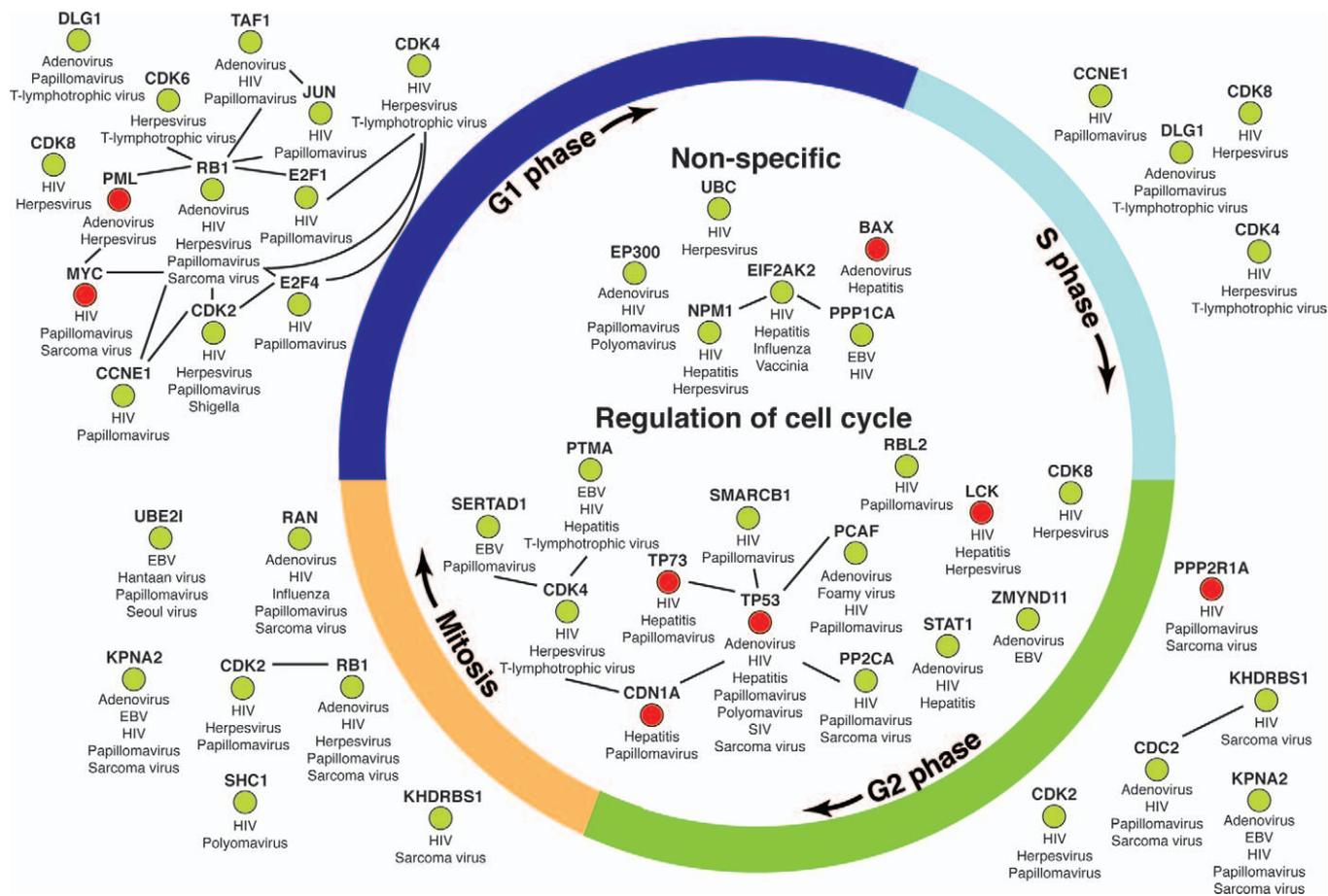
Figure 4. Human Cell Cycle Proteins Interacting with Multiple Viral Pathogen Groups

Enriched network of human proteins annotated with "cell cycle." The subset of proteins labeled as "Non-specific" are those not annotated with any function more specific than "cell cycle" in GO. If a protein participates in multiple phases, then it appears in each phase. An edge connecting two proteins denotes a known interaction in the human PPI network. Human proteins highlighted in red are those known to be involved in the induction of apoptosis.

doi:10.1371/journal.ppat.0040032.g004

TAT protein interacts with the human RB1 protein to manipulate normal cell cycle conditions and promote viral gene expression. The HIV long terminal repeat (LTR) is responsible for integrating viral DNA into the host genome and also acts as a promoter and enhancer of viral proteins. The LTR is most active in the early G_1 phase and the activity of the LTR diminishes as the cell progresses through the G_1 phase and enters the S phase [48]. Therefore, the extension of the G_1 phase may increase activity of the LTR and the eventual production of more viral proteins. In the case of Papillomavirus, the VE6 protein in Papillomavirus has been shown to manipulate the cell cycle by altering mitotic checkpoint fidelity through its effect on CDC2 activity and inactivation of TP53 [55]; it interacts with ten human proteins in Figure 4.

The human DLG1 protein is a "discs large homolog" that is essential for the transition from the G_1 to S phase of the cell cycle. This protein interacts with three pathogens: Adenovirus, Papillomavirus, and T-lymphotrophic virus [56,57]. The direct interaction of Papillomavirus proteins with human DLG1 has been implicated in development of HPV-related cancer [58].

Our analysis also identifies a network of human proteins enriched with the GO function "transcription regulator activity" (p-value 3.22×10^{-7}, rank 15/89) (see supplementary website for image). The portion of Figure 4 corresponding to the G_1 phase includes the transcription factors E2F1, E2F4, and TAF1. Each of these proteins plays a key role in normal cell cycle progression from G_1 to S phase. E2F1 and E2F2 interact with two pathogens, HIV and Papillomavirus [48,59,60]. TAF1 interacts with three pathogens, Adenovirus, HIV, and Papillomavirus [61–63]. By blocking the interaction of RB1 and various transcription factors, viral pathogens are able to prevent the cell from advancing into the S phase. This event extends the G_1 phase of the cell cycle and allows the transcription of viral genetic material.

Regulate apoptosis. An important step in viral pathogenesis is the regulation of host cell apoptosis. During the initial process of infection, prevention of apoptosis is important to allow the replication of viral genetic material. However, promotion of apoptosis has been implicated in the progression of infection. Our results underscore both phenomena. Several host proteins involved in the control of cellular apoptosis are targeted by viral pathogens (human proteins highlighted in red in Figure 4). One of the key regulators of apoptosis, and perhaps the most studied human
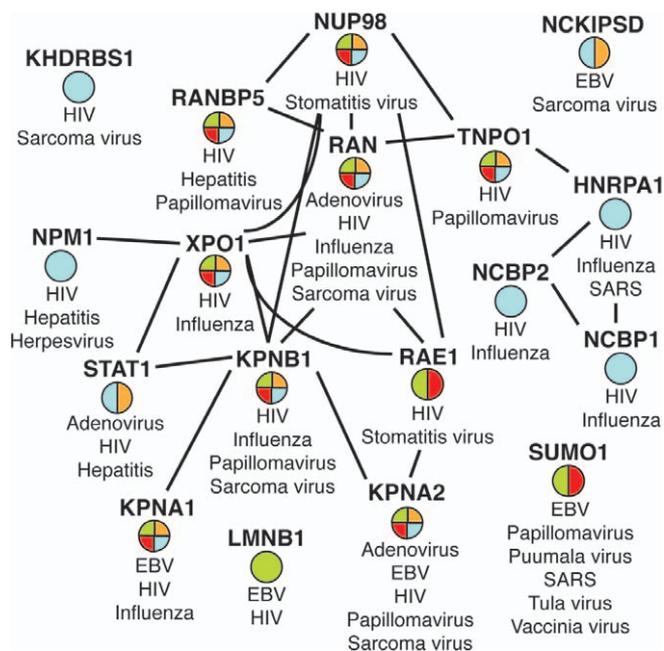
Figure 5. Human Nuclear Membrane Proteins Interacting with Multiple Viral Pathogen Groups

Enriched network of human proteins annotated with "nuclear transport" (blue), "nuclear membrane part" (green), "protein import" (orange), and "nuclear pore" (red). An edge connecting two proteins denotes a known interaction in the human PPI network.
doi:10.1371/journal.ppat.0040032.g005

protein, is TP53. TP53 interacts with seven viral pathogens: Adenovirus, Hepatitis, HIV, Papillomavirus, Polyomavirus, Sarcoma virus, and Simian virus [20, 64–70]. Interactions with Adenovirus, Hepatitis, and Papillomavirus are responsible for preventing apoptosis of the infected human cell. Adenovirus E1B and E4 proteins bind with and inactivate TP53 [71,72]. The human Survivin protein is an apoptosis inhibitor that is repressed by TP53 [73]. The repression of Survivin is necessary for the human cell to activate apoptotic programming. Another study shows that the HIV VPR protein can directly upregulate the human Survivin protein [74]. These studies suggest a common mechanism for viral inhibition of apoptosis of the host cell. TP53 interacts with a number of Hepatitis proteins including the Core protein; Core has been shown to augment TP53's transcriptional activity during infection to promote production of viral proteins and deregulate cell cycle checkpoint controls and block TP53-mediated apoptosis [75,76]. Papillomavirus VE6 interacts with human TP53 to promote degradation of TP53 and prevent apoptotic programming of the infected cell [77]. In contrast to these phenomena, the viral HIV protein TAT has been shown to assist in the progression of HIV infection by attaching to uninfected host T cells and triggering cell death via apoptosis [78,79].

Transport viral material across the nuclear membrane. Since viruses lack the machinery needed to replicate their genomes, viral genetic material must first cross the barrier from the cytoplasm into the nucleus in order to make use of the host's transcriptional machinery. Our analysis identifies a subset of human proteins enriched in four GO functions

related to this important step: "nuclear transport" (p-value 2.32×10^{-5}, rank 24/89), "nuclear membrane part" (p-value 5.61×10^{-5}, rank 28/89), "protein import" (p-value 0.001, rank 41/89), and "nuclear pore" (p-value 0.018, rank 69/89). Figure 5 displays this network. The layout in Figure 1 displays these proteins both in the region labeled "cytoplasm" and in the region labeled "nucleus."

The nuclear pore is a large protein complex that spans the nuclear membrane and allows for the transport of molecules across the nuclear envelope including proteins and RNA. There are ten human proteins that are part of the nuclear pore and targeted by multiple pathogens. These are the nodes containing a red section in Figure 5. Although smaller molecules may freely pass through the nuclear pores of the nuclear envelope, larger macromolecules require the assistance of karyopherins. Karyopherins may act as importins or exportins. Karyopherins bind to their cargo; after they cross the nuclear envelope, an interaction with the human RAN protein releases the bound partner. Figure 5 contains five human karyopherin proteins (KPNA1, KPNA2, KPNB1, RANBP5, TNPO1) as well as the human RAN protein, which interacts with five pathogens: Adenovirus, HIV, Influenza, Papillomavirus, and Sarcoma virus [20,80]. The human protein KPNB1 interacts with four pathogens: HIV, Papillomavirus, Influenza, and Simian virus [20,39,81,82]. In the case of HIV, one of the interacting partners of the human KPNB1 protein is REV. KPNB1 binds and mediates the nuclear import of the HIV REV protein. Once inside the nucleus, REV binds to unspliced viral mRNA and exports it from the nucleus to be translated [6]. REV is able to move between the nucleus and cytoplasm because it contains both a nuclear localization signal and a nuclear export signal. The human RANBP5 protein interacts with three pathogens: HIV, Hepatitis, and Papillomavirus [83–85]. The Hepatitis interactor for RANBP5 is the viral 5A protein. While little is known about the RANBP5 protein, studies suggest that the viral 5A protein may interact with RANBP5 and block secretion of cytokines produced in response to a viral infection [83]. This network highlights the ability of viral pathogens to make use of host machinery in order to translate their own genetic material and at the same time prevent the activation of a viral immune response.

Human Proteins Targeted by Multiple Bacterial Pathogens

Although the number of human-bacteria PPIs gathered in this study is small (only 174), our methods identified an important subset of human proteins enriched for functions involved in immune response and interacting with multiple bacterial pathogen groups. Figure 6 displays a subset of the multibacterial set that is enriched in four GO functions: "immune system process" (p-value 1.397×10^{-9}, rank 1/28), "response to wounding" (p-value 3.93×10^{-4}, rank 8/28), "immune response" (p-value 0.002, rank 14/28), and "I-κB kinase/NF-κB cascade" (p-value 0.012, rank 18/28). The proteins contained in this image are located in the top-right corner of Figure 2.

These functions are tied together by the Toll-Like Receptors (TLRs) and the protein IRAK1 found in the network in Figure 6. TLRs are a special class of cell-surface proteins that play a role in recognizing the presence of a pathogen and activating an immune response against the pathogen. The TLR/IRAK complex stimulates the activity of
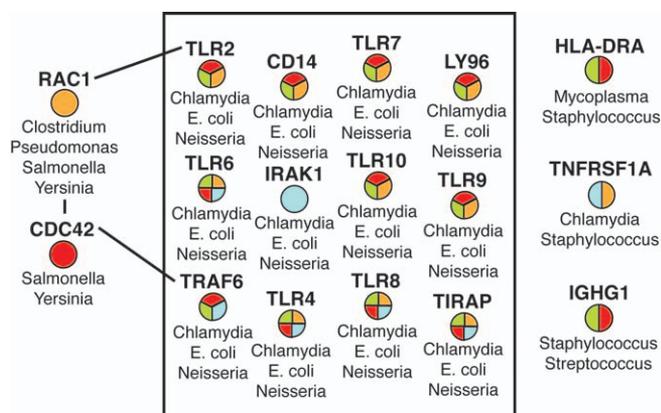
Figure 6. Human Immune System Proteins Interacting with Multiple Bacterial Pathogen Groups

Enriched network of human proteins annotated with "immune system process" (red), "response to wounding" (orange), "immune response" (green), and "I-κB kinase/NF-κB cascade" (blue). The proteins in the black box form a dense network of PPIs; we have left these out for clarity. An edge connecting two proteins denotes a known interaction in the human PPI network.

doi:10.1371/journal.ppat.0040032.g006

NF-κB [86–88], a complex of proteins that act as a TF for activating the production of a set of proteins in response to stimuli such as stress, cytokines, and bacterial or viral antigens.

The human TLRs and IRAK1 protein interact with the pathogen proteins FLIC (*E. coli*), HSP60 (Chlamydia), and PIB (Neisseria) [20]. FLIC is a flagellin protein. TLR4 and TLR5 contain a specific innate immune receptor for recognizing bacterial flagella [5,89]. HSP60 is a heat-shock protein that stimulates an immune response via TLR2 and TLR4 [90]. PIB is an outer membrane protein that is known to be recognized by TLR2, TLR4, and TLR9 [7].

Another human protein included in this network is HLA-DRA, which is part of the major histocompatibility complex (MHC). The MHC plays an important role in the immune system. HLA-DRA belongs to the class II MHC; proteins in this class belong to the lysosomal compartment of the cell, which contains digestive enzymes that kill engulfed foreign particles such as viruses or bacteria. The two bacterial partners for HLA-DRA are Mycoplasma and Staphylococcus [91,92]. In the case of Mycoplasma, the interacting partner is the MAM superantigen, which is known to contribute to autoimmune disease by activating proinflammatory mono-kines such as interleukin 1β and the tumor necrosis factor α [93].

Other Highly Targeted Human Proteins

The networks in Figures 1 and 2 contain a number of other human proteins targeted by more than two pathogen groups. We discuss two of these proteins—STAT1 and EP300.

Viral pathogens also interact with other human proteins involved in immune response pathways that are not included in the network in Figure 6. An example is the human protein STAT1. When the cell recognizes the presence of foreign material, it activates an immune response as a defense mechanism to either remove the foreign material or cause the cell to undergo apoptosis. During this process, STAT1 is tyrosine- and serine-phosphorylated and forms a homodimer

known as IFN-γ-activated factor (GAF). GAF migrates to the nucleus where it binds to specific *cis*-elements to drive the cell to produce interferons, agents that inhibit viral replication within other cells of the body [94]. STAT1 interacts with Adenovirus, HIV, and Hepatitis [95–97]. Hepatitis POLG is part of the pathogen core complex that allows the virus to avert host antiviral response by binding to host STAT1 and inhibiting its activity [98].

Within the nucleus, we see pathogens target the human protein EP300, a histone acetyltrans-ferase that regulates transcription via chromatin remodeling. EP300 interacts with Adenovirus, HIV, Papillomavirus, and Polyomavirus [99–102]. The pathogen Adenovirus targets human EP300 via E1A. E1A is an oncoprotein that stimulates cell growth and inhibits differentiation by binding to the EP300/CBP complex and deregulating cellular transcription programs [103]. Papillo-mavirus protein VE7 shares many functional and structural similarities with E1A and is an interacting partner of human EP300. The disruption of normal growth conditions brought about by the E1A-EP300 interaction leads to the development of cervical cancer [104]. In the case of HIV, the viral TAT protein targets human EP300. The resulting complex regulates TAT transactivating activity and may assist in the integration of viral genetic material into human DNA [105].

Conclusions

We have provided a general overview of the landscape of human proteins interacting with pathogens and demonstrated that pathogens preferentially interact with two classes of human proteins: hubs (i.e., proteins that interact with many other human proteins) and bottlenecks (i.e., proteins that lie on many shortest paths) in the human PPI network. We identified GO functions over-represented in human proteins interacting with pathogens. Biclustering analysis demonstrated that many sets of pathogen groups target the same processes in the human cell, even if they interact with different proteins.

We constructed networks of PPIs between human proteins that interact with at least two viral pathogen groups and with at least two bacterial pathogen groups. Consideration of the GO functions enriched in these networks provided insights into numerous pathways targeted or triggered by multiple pathogens: control and deregulation of the cell cycle; import of pathogen proteins into the nucleus in an attempt to subvert the host's DNA replication and transcription machinery; manipulation of host cellular programs such as apoptosis; immune response and activation of NF-κB pathways via the TLR/IRAK complex.

A striking aspect of this network is that human proteins that mediate pathogen effects are often proteins in cancer pathways (e.g., RB1, TP53, and STAT1). We note that only some of the pathogens targeting such proteins are known to cause cancer themselves (e.g., Herpesvirus and Papillomavirus). In fact, a number of parallels are becoming evident between infection and cancer; for instance, in the part that TLRs play in angiogenesis and their potential as targets for therapeutics [106,107] and the role that viruses may play in the development of inflammatory diseases and cancer [108]. Cell cycle regulators and many TFs have been extensively studied in the context of mediating tumor formation. Our observation that they are also communication vehicles for

Table 2. Interaction Method and Support Summary

Support	Method	Interaction Count	Fraction of Interactions
Experimental methods	Reactome curated	7,229	0.69
	Not specified	2,305	0.69
	Yeast two-hybrid	419	0.22
	Pull down	314	0.04
	Coimmunoprecipitation	210	0.03
	Other technology	210	0.02
	Observed by multiple methods	260	0.02
Literature support	Described in one paper	9,810	0.94
	Described in two papers	198	0.02
	Described in more than two papers	469	0.04

Summary of experimental methods and literature support for the host-pathogen PPIs in our dataset. The experimental group "Not specified" denotes that there was no observation method listed in any database. Note that every interaction in this study must have at least one piece of literature support.
doi:10.1371/journal.ppat.0040032.t002

pathogens suggests that the link between pathogen infection and cancer may be worthy of further experimental studies.

An important outcome of such a comparative study is the identification of human proteins to target experimentally for developing therapeutics. We provide a file on the supplementary website that contains the degree, centrality, the number of pathogen interactors, and the most specific annotations in each of the three GO hierarchies for each human protein that interacts with at least one pathogen protein. We provide this data as a resource for researchers interested in prioritizing antiviral and antibacterial targets.

We reiterate that our results should be interpreted with caution since no single pathogen may target all the proteins we analyze. As interactions between host and pathogen molecules are discovered on genome-wide scales [109], computational analyses such as those presented in this paper may provide a more detailed understanding of the landscape of host pathways and processes that pathogens target.

Methods

Datasets used. We downloaded all datasets used in this study in August 2007. We gathered 10,477 experimentally detected and manually curated protein-protein interactions (PPIs) between human and pathogen proteins and 75,457 experimentally verified PPIs between human proteins from primary literature [109] and seven databases: the Biomolecular Interaction Network Database [21], the Database of Interacting Proteins [19], the Human Protein Reference Database [23], IntAct [18], the Molecular INTeraction database [17], the Munich Information Center for Protein Sequences [22], and Reactome [20]. Table 2 contains statistics on the experimental methods that yielded these PPIs and the literature support for the PPIs. These interactions cover 190 different pathogen strains. Two pathogens—HIV and Hepatitis—account for 88.4% (9,268) of the human-pathogen PPIs. To mitigate this bias, we merged pathogen strains into 54 groups based on taxonomic similarity: each group contains pathogens belonging to the same genus, or, in the case of viruses, the same family. The 54 pathogen groups contain 35 viral, 17 bacterial, and two protozoan groups. We constructed lists of unique human proteins interacting with each group. Table 3 summarizes the number of interactions acquired for each pathogen group. For some analyses, we consider a human PPI network assembled from unbiased high-throughput experiments [14,15,37] and a network constructed from only manually curated human PPIs [20,23]. These networks contain 13,324 and 59,396 interactions, respectively. We obtained functional annotations from the Gene Ontology (GO) [26].

Notation. We represent the set of known interactions between human proteins as an undirected graph $G(V, E)$, where V is the set of nodes (proteins) and E is the set of edges (interactions). Let M be the set of pathogen groups. We say that a pathogen group P interacts with a human protein s if s interacts with a protein in P. For a pathogen group $P \in M$, we define $V_P \subseteq V$ to be the set of human proteins that interact with P. Let $T = \cup_{P \in M}$ be the set of proteins that interact with at least one pathogen. Let T_V (respectively, T_B) be the set of human proteins that interact with at least one viral (respectively, one bacterial) group. Let $T^{(k)}_V \subseteq T_V$ (respectively, $T^{(k)}_B \subseteq T_B$) be the set of human proteins that interact with at least k viral (respectively, k bacterial) pathogen groups; by definition, $T^{(1)}_V \equiv T_V$ and $T^{(1)}_B \equiv T_B$. We now describe in detail the tests we use to analyze T_B, T_V, $T^{(2)}_B$, $T^{(2)}_V$, and the 54 V_P sets.

Analysis of degree in the human PPI network. The *degree* of a protein in a graph is the number of interactions in which it participates, not including self-interactions. We plot distributions of the degrees of four sets of proteins in G: (i) V, the set of all proteins in G; (ii) T_B, the set of all human proteins interacting with at least one bacterial pathogen group; (iii) T_V, the set of all human proteins interacting with at least one viral pathogen group; and (iv) $T^{(2)}_V$, the set of human proteins interacting with at least two viral pathogen groups. In this analysis, we ignore $T^{(2)}_B$ since it contains only 20 proteins. If the distributions of T_B and T_V are more biased towards high degree proteins than the distribution for V, then we hypothesize that viral and bacterial pathogens have evolved to interact with hub proteins in the human PPI network.

Analysis of betweenness centrality in the human PPI network. The degree of a protein captures only its local connectivity. Centrality captures both global and local features of a protein's importance in a network. In this paper, we use the notion of a protein's betweenness centrality [110]. A protein with high betweenness centrality is characteristic of a bottleneck in an interaction network (i.e., there are many paths that pass through this protein) [34].

We define the *betweenness centrality* $bc(v)$ of a protein v as the fraction of shortest paths in G between all protein pairs (u,w) that pass through the protein v. Given $u, v, w \in V$, let σ_{uw} denote the number of shortest paths between proteins u and w. There may be multiple equally long paths between u and w that are shorter than any other path between u and w. Let $\sigma_{uw}(v)$ denote the number of these that pass through v. Then the betweenness centrality of v is

$$bc(v) = \sum_{\substack{u, w \in V \\ u, w \neq v}} \frac{\sigma_{uw}(v)}{\sigma_{uw}}$$

In our analysis, we divide $bc(v)$ by the number of pairs of nodes in G, yielding a quantity between 0 and 1. We use the algorithm devised by Brandes [111] to compute the betweenness centrality of all nodes in G. This algorithm runs in time proportional to the product of the number of nodes in G and the number of edges in G. As with the degree analysis, we plot distributions of the betweenness centrality for V, T_B, T_V, and $T^{(2)}_V$. If the distributions for T_B, T_V, and $T^{(2)}_V$ are biased toward higher values of centrality than the distribution for V, we hypothesize that pathogens have evolved to interact with bottlenecks in the human PPI network.

Gene set enrichment analysis. Let L be the ranked list of the proteins in V, where we rank the proteins either by degree or by betweenness centrality. Given L and a predefined set S of proteins of

Table 3. Interaction Summary

Group	Number of PPIs	Number of Strains	Number of Unique Targeted Human Proteins	Number of Targeted Proteins in Human PPI Network
HIV	8,024	44	743	671
Hepatitis	1,244	16	109	93
Influenza	287	4	76	76
Papillomavirus	229	12	96	94
Epstein-Bar virus	211	2	135	121
Adenovirus	80	9	60	59
Herpesvirus	64	20	54	54
Yersinia	57	3	56	45
Sarcoma virus	52	6	36	35
T-lymphotrophic virus	25	2	23	23
E. coli	22	1	20	20
Chlamydia	19	1	19	19
Neisseria	16	1	16	16
Streptococcus	14	5	8	8
Vaccinia virus	13	4	7	7
Staphylococcus	12	3	10	7
Pseudomonas	11	1	9	9
Measles virus	10	3	4	4
Polyomavirus	8	3	8	8
Leukemia virus	7	1	7	6
Shigella	5	1	4	4
Anemia virus	4	4	2	1
Bacillus	4	3	4	4
Hantaan virus	4	1	4	4
SARS	4	1	4	4
Clostridium	3	3	2	2
Dengue virus	3	3	3	3
Rotavirus	3	3	2	2
Echovirus	3	2	1	1
Helicobacter	3	2	2	2
Salmonella	3	1	3	3
Seoul virus	3	1	3	3
Listeria	3	1	2	2
SIV	2	2	2	2
Orf virus	2	2	1	1
Foamy virus	2	1	2	2
Puumala virus	2	1	2	2
Stomatitis virus	2	1	2	2
Mycoplasma	2	1	2	1
Sendai virus	1	1	1	1
Nucleopolyhedrovirus	1	1	1	1
Rabies virus	1	1	1	1
Toxoplasma	1	1	1	0
Poliovirus	1	1	1	1
Nipah virus	1	1	1	1
Klebsiella	1	1	1	1
Enterobacteria	1	1	1	1
Mokola virus	1	1	1	1
West Nile virus	1	1	1	1
Tula virus	1	1	1	1
Corynephage	1	1	1	1
Ebola virus	1	1	1	1
Campylobacter	1	1	1	1
Plasmodium	1	1	1	1
TOTAL	10,477	190	1,233	1,109

Overview of human-pathogen PPIs. For each pathogen group we list the total number of PPIs involving pathogen proteins in that group, the number of strains in that group, the number of unique human proteins interacting with that group, and the number of these that interact with at least one other human protein.
doi:10.1371/journal.ppat.0040032.t003

interest (e.g., those interacting with HIV), we use GSEA to determine whether the proteins contained in S are randomly distributed throughout L or concentrated at the top. In the ranked list L, let l_i be the value (of degree or centrality) at index i; $1 \leq i \leq |L|$. We abuse notation and say that an index i is an element of S if the protein whose rank is i belongs to S. First, we compute $m = \Sigma_{i \in L} l_i$, the sum of all the values in L. Next, for each index i in L, we compute two values:

$$P_{hit}(S, i) = \sum_{j \in S, \, j \leq i} \frac{l_j}{m}$$

Table 4. Gene List

Gene Name	UniProt ID(s)	Protein Name
5A	Q9YKI3_9HEPC	Non-structural protein 5A
CCR5	CCR5_HUMAN	C-C chemokine receptor type 5
CD4	CD_HUMAN	T cell surface glycoprotein CD4 precursor
CDCs	CDC[2,4,6,8]_HUMAN	Cell division control protein
CDK2	CDK2_HUMAN	Cell division protein kinase 2
CREBBP	CBP_HUMAN	CREB-binding protein
CXCR4	CXCR4_HUMAN	C-X-C chemokine receptor type 4
DLG1	DLG1_HUMAN	Disks large homolog 1
E1A	E1A_ADE02	Early E1A 32 kDa protein
E1B	E1BL_ADE02	E1B protein, large T-antigen
E2Fs	E2F[1,2]_HUMAN	Transcription factor E2F
E4	E4OR1_ADE09	E4-ORG1
EP300	EP300_HUMAN	Histone acetyltransferase p300
FLIC	FLIC_ECOLI	Flagellin
GTF2B	TF2B_HUMAN	Transcription initiation factor IIB
HDACs	HDAC[1,2]_HUMAN	Histone deacetylase
HLA-DRA	2DRA_HUMAN	HLA class II histocompatibility antigen
HSP60	O84760_CHLTR	Heat shock protein 60
KPNAs	IMA[1,2]_HUMAN	Importin subunit alpha
KPNB1	IMB1_HUMAN	Importin subunit beta-1
MAM	Q48898_MYCAT	Superantigen precursor
PIB	OMPB1_NEIMB	Major outer membrane protein PIB precursor
POLG	POLG_HCV1	Genome polyprotein
RAN	RAN_HUMAN	GTP-binding nuclear protein Ran
RANBP5	IMB3_HUMAN	Importin beta-3
RB1	RB_HUMAN	Retinoblastoma-associated protein
REV	REV_HV1H2	Rev Protein
SP1	SP1_HUMAN	Transcription factor Sp1
STAT1	STAT1_HUMAN	Signal transducer and activator of transcription 1-alpha/beta
SUMO1	SUMO1_HUMAN	Small ubiquitin-related modifier 1 precursor
TAF1	TAF1_HUMAN	Transcription initiation factor TFIID subunit 1
TAT	TAT_HV1H2	Protein Tat
TBP	TBP_HUMAN	TATA-box-binding protein
TLRs	TLR[2,4,5,9]_HUMAN	Toll-like receptor precursor
TNPO1	TNPO_HUMAN	Transportin-1
TP53	P53_HUMAN	Cellular tumor antigen p53
VE6	VE6_HPV11	Protein E6
VE7	VE7_HPV11	Protein E7

List of genes referenced in this manuscript.
doi:10.1371/journal.ppat.0040032.t004

$$P_{miss}(S, i) = \sum_{j \notin S,\ j \leq i} \frac{1}{|L| - |S|}$$

Thus, $P_{hit}(S, i)$ measures the weighted fraction of proteins with index at most i that are in S and $P_{miss}(S, i)$ measures the fraction of proteins with index at most i that are not in S. We handle multiple ranks with identical values by computing these two values only at the largest rank for each unique value in L. Finally, we define the enrichment score as the largest positive value of $P_{hit}(S, i)$ - $P_{miss}(S, i)$, i.e.,

$$es(S, L) = \max_{1 \leq i \leq |L|} \left(\max(P_{hit}(S, i) - P_{miss}(S, i), 0) \right)$$

A large positive value of $es(S, L)$ indicates that the proteins in S have high degree or high betweenness centrality. Note that our modification of the original definition of the enrichment score [35] ensures that if S mainly contains proteins with low degree or betweenness centrality, then the score will be close to 0, since $P_{hit}(S, i) - P_{miss}(S, i)$ will be negative for most indices. We record the rank i that yields $es(S, L)$; the column titled "#proteins contributing" in Table S1 of the supplementary data displays these numbers. To compute p-values for an observed enrichment scores, we generate a null distribution of scores by repeatedly selecting $|S|$ random nodes in L and computing the score for each random subset of nodes. We repeat this process 1,000,000 times and estimate the p-value for s as the fraction of random sets whose score is at least as large as s. We obtain

our results by testing each of 57 sets: T_B, T_V, $T^{(2)}_V$, and the sets V_P corresponding to each of the 54 pathogen groups.

Functional enrichment. We isolate functionally coherent subsets of human proteins among the sets T_B, T_V, $T^{(2)}_B$, $T^{(2)}_V$, and the sets V_P corresponding to each of the 54 pathogen groups using a test for functional enrichment. Given the hierarchical structure of the Gene Ontology (GO) [26], we account for dependencies between annotations by using the method proposed by Grossman et al. [112]. Let S be a set of proteins of interest (e.g., the set of proteins interacting with HIV). We aim to compute GO functions that annotate a surprisingly large number of proteins in S. To this end, for each function f in GO, we count s_f, the number of proteins in S annotated with f and $s_{pa(f)}$, the number of proteins in S annotated by at least one parent of f. We also compute v_f and $v_{pa(f)}$, the number of proteins in V annotated by f and by at least one parent of f, respectively. With these four counts in hand, we use the hypergeometric distribution to compute the probability $p_f(S, V)$ of drawing s_f or more proteins from a set of v_f marked proteins when we select $s_{pa(f)}$ proteins at random from a universe of $v_{pa(f)}$ proteins:

$$p_f(S, V) = \sum_{k=s_f}^{\min(s_{pa(f)}, v_f)} \frac{\binom{v_f}{k} \binom{v_{pa(f)} - v_f}{s_{pa(f)} - k}}{\binom{v_{pa(f)}}{s_{pa(f)}}}$$

We account for multiple hypothesis testing using the method of Benjamini and Hochberg [113]. We consider only functions enriched

with a *p*-value of at most 0.05. Note that different enriched functions may annotate identical sets of human proteins. In each such case, we group the functions and associate the most enriched function (and its *p*-value) with the group. To report enrichment ranks, we sort the groups in increasing order of *p*-value. Although not discussed in this paper we repeat this analysis using T (rather than V) as the universe of proteins. With T as the universe, we expect to find functions that distinguish between the pathogens. The results with T as universe are available on our supplementary website.

Biclustering of enriched functions. We compute enriched functions in each of the 54 sets of human proteins interacting with each pathogen group. We construct a binary matrix whose rows are enriched functions and whose columns are pathogen groups. An entry is one in this matrix if and only if the function is enriched with a *p*-value of at most 0.05 in the pathogen. In this binary matrix, we define a bicluster to be a subset R of rows and a subset C of columns such that each row-column pair in $R \times C$ contains a one. We also require a bicluster to be closed, i.e., each row not in R (respectively, column not in C) contains a zero in at least one column in C (respectively, row in R). We use the Bimax algorithm to compute all closed biclusters in this binary matrix [114].

Supporting Information

Figure S1. Protein Degree–Centrality Scatter-Plots

Log-log scatter-plots of each protein contained within the three networks used in this study: (A) the whole human PPI network (11,463 proteins), (B) the high-throughput human PPI network (4,986 proteins), and (C) the manually curated human PPI network (8,704 proteins). The *x*-axis is the degree and the *y*-axis is the centrality of a protein within its respective network.

Found at doi:10.1371/journal.ppat.0040032.sg001 (2.5 MB TIF).

Table S1. Relative Node Occurrences

Relative occurrences of four types of nodes in each of the three networks: Whole human PPI network (W), the human PPI network yielded by High-Throughput experiments (HT), and the human PPI network consisting only of Manually Curated PPIs (MC). The "Fraction" column defines the cutoff at which a protein is considered a hub or a bottleneck. The other columns represent the fraction of hub-bottleneck, non-hub-bottleneck, hub-non-bottleneck, and non-hub-non-bottleneck proteins in the network using that cutoff.

Found at doi:10.1371/journal.ppat.0040032.st001 (46 KB PDF).

Table S2. Detailed GSEA Results

Summary of GSEA results with and without human-HIV interactions for three networks: Whole human PPI network (W), the human PPI network yielded by High-Throughput experiments (HT), and the human PPI network consisting only of Manually Curated PPIs (MC). We report *p*-values only for the sets of human proteins in Figure 3. The "#proteins in group" column displays the total number of

human proteins in that group. The "ES" column displays the enrichment score calculated by GSEA. The column titled "#proteins contributing" displays the number of proteins contributing to the ES score (see Methods for details.) The column titled "Jaccard's" lists the Jaccard coefficient between the two sets of proteins contributing to the ES score for degree and for centrality.

Found at doi:10.1371/journal.ppat.0040032.st002 (67 KB PDF).

Table S3. Detailed GSEA Results for Individual Pathogen Groups

Summary of GSEA results for individual pathogen groups for three networks: Whole human PPI network (W), the human PPI network yielded by High-Throughput experiments (HT), and the human PPI network consisting only of Manually Curated PPIs (MC). We report *p*-values only for the sets of human proteins in Figure 3. The "#proteins in group" column displays the total number of human proteins in that group. The "ES" column displays the enrichment score calculated by GSEA. The column titled "#proteins contributing" displays the number of proteins contributing to the ES score (see Methods for details.) The column titled "Jaccard's" lists the Jaccard coefficient between the two sets of proteins contributing to the ES score for degree and for centrality. We only report groups that are enriched for both degree and centrality.

Found at doi:10.1371/journal.ppat.0040032.st003 (72 KB PDF).

Text S1. Supporting Text

Testing correlation between protein degree and protein centrality for GSEA analysis and GSEA analysis for individual pathogen groups.

Found at doi:10.1371/journal.ppat.0040032.sd001 (59 KB PDF).

Accession Numbers

Table 4 contains a list of all the proteins discussed in this paper and their corresponding UniProt ids and descriptions.

Acknowledgments

References

1. Huang L, Bosch I, Hofmann W, Sodroski J, Pardee AB (1998) Tat protein induces human immunodeficiency virus type 1 (HIV-1) coreceptors and promotes infection with both macrophage-tropic and T-lymphotropic HIV-1 strains. J Virol 72: 8952–8960.
2. LaCount DJ, Vignali M, Chettier R, Phansalkar A, Bell R, et al. (2005) A protein interaction network of the malaria parasite *Plasmodium falciparum*. Nature 438: 103–107.
3. Filippova M, Parkhurst L, Duerksen-Hughes PJ (2004) The human papillomavirus 16 E6 protein binds to Fas-associated death domain and protects cells from Fas-triggered apoptosis. J Biol Chem 279: 25729–25744.
4. Geisberg JV, Lee WS, Berk AJ, Ricciardi RP (1994) The zinc finger region of the adenovirus E1A transactivating domain complexes with the TATA box binding protein. Proc Natl Acad Sci U S A 91: 2488–2492.
5. Hayashi F, Smith KD, Ozinsky A, Hawn TR, Yi EC, et al. (2001) The innate immune response to bacterial flagellin is mediated by toll-like receptor 5. Nature 410: 1099–1103.
6. Henderson BR, Percipalle P (1997) Interactions between HIV Rev and nuclear import and export factors: the Rev nuclear localisation signal mediates specific binding to human importin-beta. J Mol Biol 274: 693–707.
7. Mogensen TH, Paludan SR, Kilian M, Ostergaard L (2006) Live *Streptococcus pneumoniae*, *Haemophilus influenzae*, and *Neisseria meningitidis* activate the inflammatory response through toll-like receptors 2, 4, and 9 in species-specific patterns. J Leukoc Biol 80: 267–277.
8. Ito T, Tashiro K, Muta S, Ozawa R, Chiba T, et al. (2000) Toward a protein-

protein interaction map of the budding yeast: a comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins. Proc Natl Acad Sci U S A 97: 1143–1147.
9. Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, et al. (2002) Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. Nature 415: 180–183.
10. Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, et al. (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature 415: 141–147.
11. Giot L, Bader JS, Brouwer C, Chaudhuri A, Kuang B, et al. (2003) A protein interaction map of *Drosophila melanogaster*. Science 302: 1727–1736.
12. Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, et al. (2001) A comprehensive two-hybrid analysis to explore the yeast protein inter-actome. Proc Natl Acad Sci U S A 98: 4569–4574.
13. Li S, Armstrong CM, Bertin N, Ge H, Milstein S, et al. (2004) A map of the interactome network of the metazoan *C. elegans*. Science 303: 540–543.
14. Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, et al. (2005) Toward a proteome-scale map of the human protein-protein interaction network. Nature 437: 1173–1178.
15. Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, et al. (2005) A human protein-protein interaction network: a resource for annotating the proteome. Cell 122: 957–968.
16. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, et al. (2000) A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. Nature 403: 623–627.
17. Zanzoni A, Montecchi-Palazzi L, Quondam M, Ausiello G, Helmer-

Citterich M, et al. (2002) Mint: a molecular INTeraction database. FEBS Lett 513: 135–140.

18. Hermjakob H, Montecchi-Palazzi L, Lewington C, Mudali S, Kerrien S, et al. (2004) IntAct: an open source molecular interaction database. Nucleic Acids Res 32: D452–D455.

19. Salwinski L, Miller CS, Smith AJ, Pettit FK, Bowie JU, et al. (2004) The database of interacting proteins: 2004 update. Nucleic Acids Res 32: D449–D451.

20. Joshi-Tope G, Gillespie M, Vastrik I, D'Eustachio P, Schmidt E, et al. (2005) REACTOME: a knowledgebase of biological pathways. Nucleic Acids Res 33: D428–D432.

21. Gilbert D (2005) Biomolecular interaction network database. Brief Bioinform 6: 194–198.

22. Guldener U, Munsterkotter M, Oesterheld M, Pagel P, Ruepp A, et al. (2006) Mpact: the MIPS protein interaction resource on yeast. Nucleic Acids Res 34: D436–D441.

23. Mishra GR, Suresh M, Kumaran K, Kannabiran N, Suresh S, et al. (2006) Human protein reference database–2006 update. Nucleic Acids Res 34: D411–D444.

24. Barsky A, Gardy JL, Hancock REW, Munzner T (2007) Cerebral: a Cytoscape plugin for layout of and interaction with biological networks using subcellular localization annotation. Bioinformatics 23: 1040–1042.

25. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13: 2498–2504.

26. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. (2000) Gene Ontology: tool for the unification of biology. The Gene Ontology consortium. Nat Genet 25: 25–29.

27. Barabasi AL, Albert R (1999) Emergence of scaling in random networks. Science 286: 509–512.

28. Almaas E (2007) Biological impacts and context of network theory. J Exp Biol 210: 1548–1558.

29. Albert R, Jeong H, Barabasi AL (2000) Error and attack tolerance of complex networks. Nature 406: 378–382.

30. Li D, Li J, Ouyang S, Wang J, Wu S, et al. (2006) Protein interaction networks of *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, and *Drosophila melanogaster*: large-scale organization and robustness. Proteomics 6: 456–461.

31. Han JDJ, Dupuy D, Bertin N, Cusick ME, Vidal M (2005) Effect of sampling on topology predictions of protein-protein interaction networks. Nat Biotechnol 23: 839–844.

32. Rachlin J, Cohen DD, Cantor C, Kasif S (2006) Biological context networks: a mosaic view of the interactome. Mol Syst Biol 2: 66.

33. Stumpf MPH, Wiuf C, May RM (2005) Subnets of scale-free networks are not scale-free: sampling properties of networks. Proc Natl Acad Sci U S A 102: 4221–4224.

34. Yu H, Kim PM, Sprecher E, Trifonov V, Gerstein M (2007) The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. PLoS Computational Biology 3: e59. doi:10.1371/journal.pcbi.0030059

35. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, et al. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102: 15545–15550.

36. Hoyle DC, Rattray M, Jupp R, Brass A (2002) Making sense of microarray data distributions. Bioinformatics 18: 576–584.

37. Ewing RM, Chu P, Elisma F, Li H, Taylor P, et al. (2007) Large-scale mapping of human protein-protein interactions by mass spectrometry. Mol Syst Biol 3: 89.

38. Ao Z, Huang G, Yao H, Xu Z, Labine M, et al. (2007) Interaction of human immunodeficiency virus type 1 Integrase with cellular nuclear import receptor Importin 7 and its impact on viral replication. J Biol Chem 282: 13456–13467.

39. Darshan MS, Lucchi J, Harding E, Moroianu J (2004) The L2 minor capsid protein of human papillomavirus type 16 interacts with a network of nuclear import receptors. J Virol 78: 12179–12188.

40. Gallay P, Hope T, Chin D, Trono D (1997) HIV-1 infection of nondividing cells through the recognition of integrase by the importin/karyopherin pathway. Proc Natl Acad Sci U S A 94: 9825–9830.

41. Pryor MJ, Rawlinson SM, Butcher RE, Barton CL, Waterhouse TA, et al. (2007) Nuclear localization of dengue virus nonstructural protein 5 through its importin alpha/beta-recognized nuclear localization sequences is integral to viral infection. Traffic 8: 795–807.

42. Punga T, Akusjarvi G (2000) The adenovirus-2 E1B-55K protein interacts with a mSin3A/histone deacetylase 1 complex. FEBS Lett 476: 248–252.

43. Brehm A, Nielsen SJ, Miska EA, McCance DJ, Reid JL, et al. (1999) The E7 oncoprotein associates with Mi2 and histone deacetylase activity to promote cell growth. EMBO J 18: 2449–2458.

44. Meier A, Alter G, Frahm N, Sidhu H, Li B, et al. (2007) MyD88-dependent immune activation mediated by HIV-1-encoded TLR ligands. J Virol 81: 8180–8191.

45. Bartz SR, Rogel ME, Emerman M (1996) Human immunodeficiency virus type 1 cell cycle control: Vpr is cytostatic and mediates G2 accumulation by a mechanism which differs from DNA damage checkpoint control. J Virol 70: 2324–2331.

46. Sinclair AJ, Fenton M, Delikat S (1998) Interactions between Epstein-Barr virus and the cell cycle control machinery. Histol Histopathol 13: 461–467.

47. Southern S, Herrington C (2000) Disruption of cell cycle control by human papillomaviruses with special reference to cervical carcinoma. Int J Gynecol Cancer 10: 263–274.

48. Kundu M, Sharma S, De Luca A, Giordano A, Rappaport J, et al. (1998) HIV-1 Tat elongates the G1 phase and indirectly promotes HIV-1 gene expression in cells of glial origin. J Biol Chem 273: 8130–8136.

49. Hanahan D, Weinberg RA (2000) The hallmarks of cancer. Cell 100: 57–70.

50. Fortunato EA, Sommer MH, Yoder K, Spector DH (1997) Identification of domains within the human cytomegalovirus major immediate-early 86-kilodalton protein and the retinoblastoma protein required for physical and functional interaction with each other. J Virol 71: 8176–8185.

51. Huang S, Lee WH, Lee EY (1991) A cellular protein that competes with SV40 T antigen for binding to the retinoblastoma gene product. Nature 350: 160–162.

52. Liu F, Green MR (1994) Promoter targeting by adenovirus E1a through interaction with different cellular DNA-binding domains. Nature 368: 520–525.

53. Liu X, Clements A, Zhao K, Marmorstein R (2006) Structure of the human Papillomavirus E7 oncoprotein and its mechanism for inactivation of the retinoblastoma tumor suppressor. J Biol Chem 281: 578–586.

54. Prasad MV, Shanmugam G (1993) Retinoblastoma gene inhibits transactivation of HIV- LTR linked gene expression upon co-transfection in He La cells. Biochem Mol Biol Int 29: 57–62.

55. Thompson DA, Belinsky G, Chang TH, Jones DL, Schlegel R, et al. (1997) The human papillomavirus-16 E6 oncoprotein decreases the vigilance of mitotic checkpoints. Oncogene 15: 3025–3035.

56. Lee SS, Weiss RS, Javier RT (1997) Binding of human virus oncoproteins to hDlg/SAP97, a mammalian homolog of the Drosophila discs large tumor suppressor protein. Proc Natl Acad Sci U S A 94: 6670–6675.

57. Thomas M, Massimi P, Navarro C, Borg JP, Banks L (2005) The hScrib/Dlg apico-basal control complex is differentially targeted by HPV-16 and HPV-18 E6 proteins. Oncogene 24: 6222–6230.

58. Kiyono T, Hiraiwa A, Fujita M, Hayashi Y, Akiyama T, et al. (1997) Binding of high-risk human papillomavirus E6 oncoproteins to the human homologue of the Drosophila discs large tumor suppressor protein. Proc Natl Acad Sci U S A 94: 11612–11616.

59. Ambrosino C, Palmieri C, Puca A, Trimboli F, Schiavone M, et al. (2002) Physical and functional interaction of HIV-1 Tat with E2F-4, a transcriptional regulator of mammalian cell cycle. J Biol Chem 277: 31448–31458.

60. Huh KW, DeMasi J, Ogawa H, Nakatani Y, Howley PM, et al. (2005) Association of the human papillomavirus type 16 E7 oncoprotein with the 600-kDa retinoblastoma protein-associated factor, p600. Proc Natl Acad Sci U S A 102: 11492–11497.

61. Carrillo E, Garrido E, Gariglio P (2004) Specific in vitro interaction between papillomavirus E2 proteins and TBP-associated factors. Intervirology 47: 342–349.

62. Geisberg JV, Chen JL, Ricciardi RP (1995) Subregions of the adenovirus E1A transactivation domain target multiple components of the TFIID complex. Mol Cell Biol 15: 6283–6290.

63. Kashanchi F, Piras G, Radonovich MF, Duvall JF, Fattaey A, et al. (1994) Direct interaction of human TFIID with the HIV-1 transactivator Tat. Nature 367: 295–299.

64. Cathomen T, Weitzman MD (2000) A functional complex of adenovirus proteins E1B-55kDa and E4orf6 is necessary to modulate the expression level of p53 but not its transcriptional activity. J Virol 74: 11407–11412.

65. Lechner MS, Laimins LA (1994) Inhibition of p53 DNA binding by human papillomavirus E6 proteins. J Virol 68: 4262–4273.

66. Longo F, Marchetti MA, Castagnoli L, Battaglia PA, Gigliani F (1995) A novel approach to protein-protein interaction: complex formation between the p53 tumor suppressor and the HIV Tat proteins. Biochem Biophys Res Commun 206: 326–334.

67. Lu W, Lo SY, Chen M, Wu Kj, Fung YK (1999) Activation of p53 tumor suppressor by hepatitis C virus core protein. Virology 264: 134–141.

68. Reichelt M, Zang KD, Seifert M, Welter C, Ruffing T (1999) The yeast two-hybrid system reveals no interaction between p73 alpha and SV40 large T-antigen. Arch Virol 144: 621–626.

69. Staib C, Pesch J, Gerwig R, Gerber JK, Brehm U, et al. (1996) p53 inhibits JC virus DNA replication in vivo and interacts with JC virus large T-antigen. Virology 219: 237–246.

70. Poulin DL, Kung AL, DeCaprio JA (2004) p53 targets simian virus 40 large T antigen for acetylation by CBP. J Virol 78: 8245–8253.

71. Dobner T, Horikoshi N, Rubenwolf S, Shenk T (1996) Blockage by adenovirus E4orf6 of transcriptional activation by the p53 tumor suppressor. Science 272: 1470–1473.

72. Punga T, Akusjarvi G (2003) Adenovirus 2 E1B-55K protein relieves p53-mediated transcriptional repression of the survivin and MAP4 promoters. FEBS Lett 552: 214–218.

73. Hoffman WH, Biade S, Zilfou JT, Chen J, Murphy M (2002) Transcriptional repression of the anti-apoptotic survivin gene by wild type p53. J Biol Chem 277: 3247–3257.

74. Zhu Y, Roshal M, Li F, Blackett J, Planelles V (2003) Upregulation of survivin by HIV-1 Vpr. Apoptosis 8: 71–79.

75. Otsuka M, Kato N, Lan K, Yoshida H, Kato J, et al. (2000) Hepatitis C virus

core protein enhances p53 function through augmentation of DNA binding affinity and transcriptional ability. J Biol Chem 275: 34122–34130.

76. Wang XW, Gibson MK, Vermeulen W, Yeh H, Forrester K, et al. (1995) Abrogation of p53-induced apoptosis by the hepatitis B virus X gene. Cancer Res 55: 6012–6016.

77. Spitkovsky D, Aengeneyndt F, Braspenning J, von Knebel Doeberitz M (1996) p53-independent growth regulation of cervical cancer cells by the papillomavirus E6 oncogene. Oncogene 13: 1027–1035.

78. Campbell GR, Pasquier E, Watkins J, Bourgarel-Rey V, Peyrot V, et al. (2004) The glutamine-rich region of the HIV-1 Tat protein is involved in T cell apoptosis. J Biol Chem 279: 48197–48204.

79. Li CJ, Friedman DJ, Wang C, Metelev V, Pardee AB (1995) Induction of apoptosis in uninfected lymphocytes by HIV-1 Tat protein. Science 268: 429–431.

80. De Luca A, Mangiacasale R, Severino A, Malquori L, Baldi A, et al. (2003) E1A deregulates the centrosome cycle in a Ran GTPase-dependent manner. Cancer Res 63: 1430–1437.

81. Nakanishi A, Shum D, Morioka H, Otsuka E, Kasamatsu H (2002) Interaction of the Vp3 nuclear localization signal with the importin alpha 2/beta heterodimer directs nuclear entry of infecting simian virus 40. J Virol 76: 9368–9377.

82. Truant R, Cullen BR (1999) The arginine-rich domains present in human immunodeficiency virus type 1 Tat and Rev function as direct importin beta-dependent nuclear localization signals. Mol Cell Biol 19: 1210–1217.

83. Chung KM, Lee J, Kim JE, Song OK, Cho S, et al. (2000) Nonstructural protein 5A of hepatitis C virus inhibits the function of karyopherin beta3. J Virol 74: 5233–5241.

84. Efthymiadis A, Briggs LJ, Jans DA (1998) The HIV-1 Tat nuclear localization sequence confers novel nuclear import properties. J Biol Chem 273: 1623–1628.

85. Nelson LM, Rose RC, Moroianu J (2003) The L1 major capsid protein of human papillomavirus type 11 interacts with Kap 2 and Kap 3 nuclear import receptors. Virology 306: 162–169.

86. Koziczak-Holbro M, Joyce C, Gluck A, Kinzel B, Muller M, et al. (2007) IRAK-4 kinase activity is required for interleukin-1 (IL-1) receptor- and toll-like receptor 7-mediated signaling and gene expression. J Biol Chem 282: 13552–13560.

87. Medzhitov R, Preston-Hurlburt P, Janeway CAJ (1997) A human homologue of the Drosophila toll protein signals activation of adaptive immunity. Nature 388: 394–397.

88. Yang RB, Mark MR, Gray A, Huang A, Xie MH, et al. (1998) Toll-like receptor-2 mediates lipopolysaccharide-induced cellular signaling. Nature 395: 284–288.

89. Mizel SB, Honko AN, Moors MA, Smith PS, West AP (2003) Induction of macrophage nitric oxide production by Gram-negative flagellin involves signaling via heteromeric toll-like receptor 5/toll-like receptor 4 complexes. J Immunol 170: 6217–6223.

90. Da Costa CUP, Wantia N, Kirschning CJ, Busch DH, Rodriguez N, et al. (2004) Heat shock protein 60 from Chlamydia pneumoniae elicits an unusual set of inflammatory responses via toll-like receptor 2 and 4 in vivo. Eur J Immunol 34: 2874–2884.

91. Petersson K, Thunnissen M, Forsberg G, Walse B (2002) Crystal structure of a SEA variant in complex with MHC class II reveals the ability of SEA to crosslink MHC molecules. Structure 10: 1619–1626.

92. Zhao Y, Li Z, Drozd SJ, Guo Y, Mourad W, et al. (2004) Crystal structure of Mycoplasma arthritidis mitogen complexed with HLA-DR1 reveals a novel superantigen fold and a dimerized superantigen-MHC complex. Structure 12: 277–288.

93. al Daccak R, Mehindate K, Hebert J, Rink L, Mecheri S, et al. (1994) Mycoplasma arthritidis-derived superantigen induces proinflammatory monokine gene expression in the THP-1 human monocytic cell line. Infect Immun 62: 2409–2416.

94. Dupuis S, Jouanguy E, Al-Hajjar S, Fieschi C, Al-Mohsen IZ, et al. (2003)

95. Izmailova E, Bertley FMN, Huang Q, Makori N, Miller CJ, et al. (2003) HIV-1 Tat reprograms immature dendritic cells to express chemoattractants for activated T cells and macrophages. Nat Med 9: 191–197.

Impaired response to interferon-alpha/beta and lethal viral disease in human STAT1 deficiency. Nat Genet 33: 388–391.

96. Lin W, Kim SS, Yeung E, Kamegaya Y, Blackard JT, et al. (2006) Hepatitis C virus core protein blocks interferon signaling by interaction with the STAT1 SH2 domain. J Virol 80: 9226–9235.

97. Look DC, Roswit WT, Frick AG, Gris-Alevy Y, Dickhaus DM, et al. (1998) Direct suppression of Stat1 function during adenoviral infection. Immunity 9: 871–880.

98. Blindenbacher A, Duong FHT, Hunziker L, Stutvoet STD, Wang X, et al. (2003) Expression of hepatitis c virus proteins inhibits interferon alpha signaling in the liver of transgenic mice. Gastroenterology 124: 1465–1475.

99. Chiao C, Bader T, Stenger JE, Baldwin W, Brady J, et al. (2001) HIV type 1 Tat inhibits tumor necrosis factor-induced repression of tumor necrosis factor receptor p55 and amplifies tumor necrosis factor alpha activity in stably tat-transfected HeLa Cells. AIDS Res Hum Retroviruses 17: 1125–1132.

100. Dorsman JC, Teunisse AF, Zantema A, van der Eb AJ (1997) The adenovirus 12 E1A proteins can bind directly to proteins of the p300 transcription co-activator family, including the CREB-binding protein CBP and p300. J Gen Virol 78 (Pt 2): 423–426.

101. Muller A, Ritzkowsky A, Steger G (2002) Cooperative activation of human papillomavirus type 8 gene expression by the E2 protein and the cellular coactivator p300. J Virol 76: 11042–11053.

102. Patel D, Huang SM, Baglia LA, McCance DJ (1999) The E6 protein of human papillomavirus type 16 binds to and inhibits co-activation by CBP and p300. EMBO J 18: 5061–5072.

103. Lang SE, Hearing P (2003) The adenovirus E1A oncoprotein recruits the cellular TR-RAP/GCN5 histone acetyltransferase complex. Oncogene 22: 2836–2841.

104. Bernat A, Avvakumov N, Mymryk JS, Banks L (2003) Interaction between the HPV E7 oncoprotein and the transcriptional coactivator p300. Oncogene 22: 7871–7881.

105. Vendel AC, Lumb KJ (2003) Molecular recognition of the human coactivator CBP by the HIV-1 transcriptional activator Tat. Biochemistry 42: 910–916.

106. Cristofaro P, Opal SM (2006) Role of toll-like receptors in infection and immunity: clinical implications. Drugs 66: 15–29.

107. McInturff JE, Modlin RL, Kim J (2005) The role of toll-like receptors in the pathogenesis and treatment of dermatological disease. J Invest Dermatol 125: 1–8.

108. Soderberg-Naucler C (2006) Does cytomegalovirus play a causative role in the development of various inflammatory diseases and cancer? J Intern Med 259: 219–246.

109. Calderwood MA, Venkatesan K, Xing L, Chase MR, Vazquez A, et al. (2007) Epstein-Barr virus and virus human protein interaction maps. Proc Natl Acad Sci U S A 104: 7606–7611.

110. Freeman LC (1977) Set of measures of centrality based on betweenness. Sociometry 40: 35–41.

111. Brandes U (2001) A faster algorithm for betweenness centrality. J Math Sociol 25: 163–177.

112. Grossman S, Baur S, Robinson PN, Vingron M (2006) An improved statistic for detecting over-represented gene ontology annotations in gene sets. In: Apostolico A, editor. RECOMB 2006. Berlin: Springer-Verlag. pp. 85–98.

113. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate—a practical and powerful approach to multiple testing. J Roy Stat Soc B Met 57: 289–300.

114. Prelic A, Bleuler S, Zimmermann P, Wille A, Buhlmann P, et al. (2006) A systematic comparison and evaluation of biclustering methods for gene expression data. Bioinformatics 22: 1122–1129.