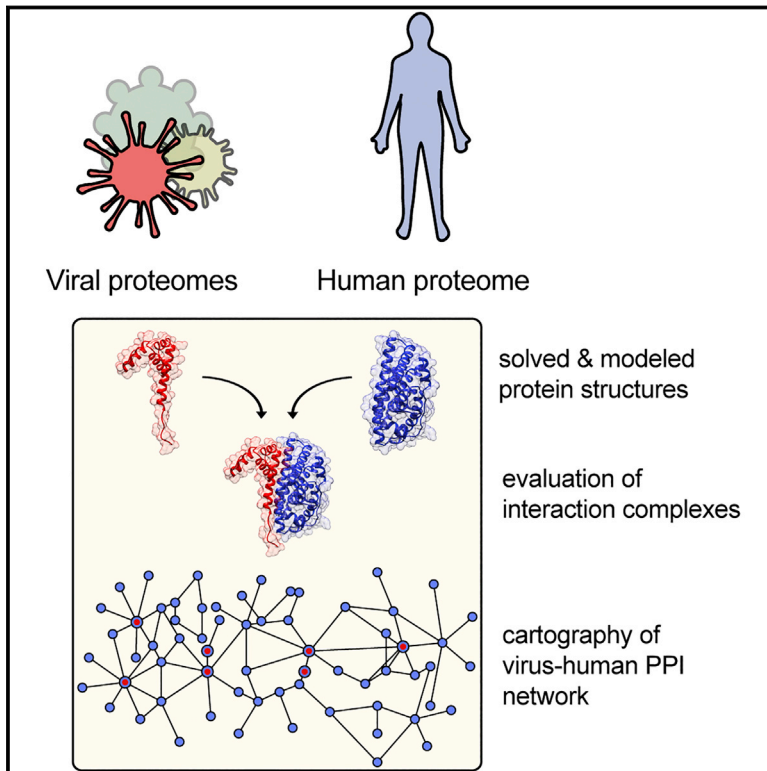


A Structure-Informed Atlas of Human-Virus Interactions

Graphical Abstract



Authors

Gorka Lasso, Sandra V. Mayer, Evandro R. Winkelmann, ..., Raul Rabadan, Barry Honig, Sagi D. Shapira

Correspondence

bh6@columbia.edu (B.H.),
ss4197@columbia.edu (S.D.S.)

In Brief

A computational approach facilitates the prediction and validation of protein-protein interactions between viruses and humans.

Highlights

- P-HIPSTer identifies ~282,000 high-confidence pan human-virus PPIs
- Cellular mediators of ZIKV infection and PPI-based classification of HPV oncogenicity
- Identification of shared/unique infection strategies employed by human viruses
- Cell-factor-centric analysis of human evolution



A Structure-Informed Atlas of Human-Virus Interactions

Gorka Lasso,^{1,2} Sandra V. Mayer,^{1,2} Evandro R. Winkelmann,^{1,2} Tim Chu,¹ Oliver Elliot,¹ Juan Angel Patino-Galindo,¹ Kernyu Park,⁴ Raul Rabadan,^{1,4} Barry Honig,^{1,3,5,6,*} and Sagi D. Shapira^{1,2,7,*}

¹Department of Systems Biology, Columbia University Medical Center, New York, NY, USA

²Department of Microbiology and Immunology, Columbia University Medical Center, New York, NY, USA

³Department of Biochemistry and Molecular Biophysics, Columbia University Medical Center, New York, NY, USA

⁴Department of Biomedical Informatics, Columbia University Medical Center, New York, NY, USA

⁵Zuckerman Mind Brain Behavior Institute, Columbia University Medical Center, New York, NY, USA

⁶Howard Hughes Medical Institute, Columbia University Medical Center, New York, NY, USA

⁷Lead Contact

*Correspondence: bh6@columbia.edu (B.H.), ss4197@columbia.edu (S.D.S.)

<https://doi.org/10.1016/j.cell.2019.08.005>

SUMMARY

While knowledge of protein-protein interactions (PPIs) is critical for understanding virus-host relationships, limitations on the scalability of high-throughput methods have hampered their identification beyond a number of well-studied viruses. Here, we implement an *in silico* computational framework (pathogen host interactome prediction using structure similarity [P-HIPSTer]) that employs structural information to predict ~282,000 pan viral-human PPIs with an experimental validation rate of ~76%. In addition to rediscovering known biology, P-HIPSTer has yielded a series of new findings: the discovery of shared and unique machinery employed across human-infecting viruses, a likely role for ZIKV-ESR1 interactions in modulating viral replication, the identification of PPIs that discriminate between human papilloma viruses (HPVs) with high and low oncogenic potential, and a structure-enabled history of evolutionary selective pressure imposed on the human proteome. Further, P-HIPSTer enables discovery of previously unappreciated cellular circuits that act on human-infecting viruses and provides insight into experimentally intractable viruses.

INTRODUCTION

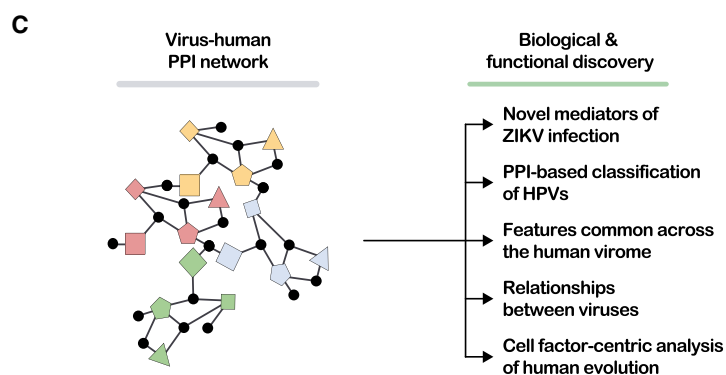
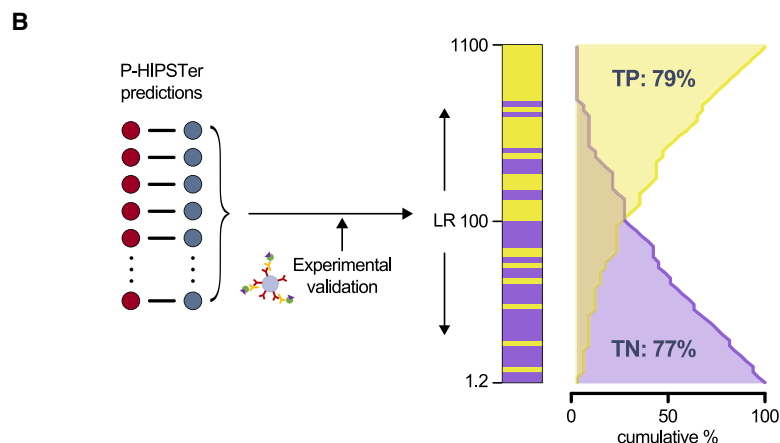
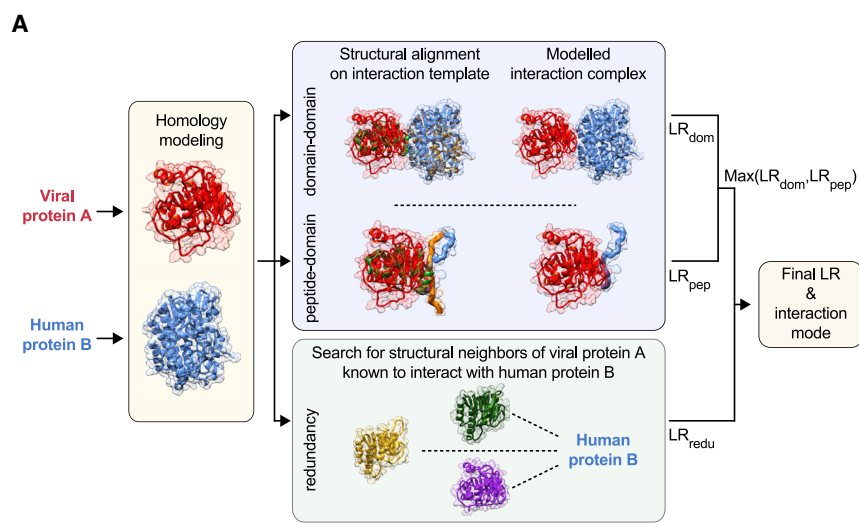
While viruses employ a complex network of protein-protein interactions (PPIs) to coopt cellular processes—such as endocytosis, transcription and capping, nuclear transport, protein translation, and secretion—host cells respond by initiating a complex transcriptional program targeted at activating innate anti-viral defenses that control viral replication and activate the adaptive immune system. In this regard, knowledge of virus-host PPIs is critical for understanding the precise series of events that control cellular responses to infection as well as mediate viral replicative

cycles. Yet, our knowledge of the machinery that mediates and controls the interaction between virus and host remains exceedingly sparse. Considerable effort involving a multitude of methods (including yeast two-hybrid assays and affinity purification) has been invested in delineating physical interactions between viral and human proteins (Nicod et al., 2017). These approaches have yielded critical insights into virus-human relationships, identification of key mediators of immunity, and discovery of cellular factors that control viral replication. However, limitations on scalability have hampered identification of PPIs en masse. Indeed, although a modest collection of PPIs exists for a handful of well-studied viruses, virtually nothing beyond genome sequence is known for the great majority of the ~1,000 human infecting viruses that have been identified and whose public health importance is unquestionable.

Here, we report a systematic interrogation of PPIs predicted for a compiled a set of 1,001 fully sequenced human infecting viruses represented by 12,237 proteins from virus-host DB, a repository of viral genomes and curated host information (Mihara et al., 2016). Based on an adaptation of the extensively validated PrePPI algorithm (predicting protein-protein interactions) (Garzón et al., 2016; Zhang et al., 2012), P-HIPSTer (pathogen-host interactome prediction using structure similarity) exploits protein structural information, taken from the Protein Data Bank (PDB) and homology modeling, to predict viral-host PPIs by accounting for both domain-domain and peptide-domain interactions (Figure 1A).

The P-HIPSTer database, comprised of ~282,000 PPIs, represents a comprehensive catalog of virus-human PPIs that spans the Baltimore classification system and is a major expansion on previously available or reported pathogen-host interactions. We have subjected *in silico* P-HIPSTer predictions to extensive statistical and empirical validation and demonstrate its ability to rediscover viral-human PPIs and cellular determinants of viral replication identified by orthogonal approaches and genome-wide screens. We report five applications of P-HIPSTer that highlight both its ability to capture known biology and to provide biological insights not available from existing tools (Figure 1C). These include: (1) validation and analysis of Zika virus (ZIKV)-human PPIs and discovery of estrogen receptor 1 (ESR1) as a major





rheostat of viral replication; (2) analysis of human papilloma virus (HPV) determinants of oncogenicity and development of an interaction-based classifier of high and low risk viruses that may be deployed in clinical settings; (3) identification of cellular pathways coopted across the human virome, many of which were not previously recognized; (4) identification of evolutionary and functional relationships across viral families and nucleic acid types that add a new layer of biological complexity to the Baltimore Classification System; and (5) discovery of selection

Figure 1. P-HIPSTER Enables Human-Virus Interactome Mapping and Interrogation

(A) P-HIPSTER uses protein structure homology modeling to evaluate viral-human PPIs mediated by domain-domain or peptide domain contacts.

(B) Empirical validation by colP of 65 predicted viral-human PPIs. Positive and negative interactions are shown in yellow and purple, respectively. TP, true positive rate; TN, true negative rate.

(C) Posterior analyses leveraging P-HIPSTER human-virus protein interactome.

See also [Figure S1](#) and [Tables S1](#) and [S2](#).

imposed by viruses on the human-primate lineage. P-HIPSTER provides a unique tool both for structure-based functional interrogation of viral proteins and for the generation of testable hypotheses that have the potential to uncover new biology. Moreover, as will be shown, the unprecedented scale of the P-HIPSTER database provides a framework for discovering shared and unique cellular pathways across 1,001 viruses representing 28 viral families known to infect humans.

RESULTS

The P-HIPSTER Algorithm, Database, and Validation

P-HIPSTER uses three sources of evidence to predict whether two proteins interact ([Figure 1A](#)). Briefly, an integrated likelihood ratio (LR) for a given PPI is based on: (1) LR that two structured domains interact based on evaluation of a model derived from a known complex comprised of their “structural neighbors,” that is, proteins with similar three dimensional folds; (2) LR that an unstructured peptide will bind to a given structured domain based on known binding motifs in the peptide sequence; and (3) LR based on evidence that multiple structural neighbors of a query protein interact with a target protein. The three LRs are then combined to yield a final LR score representing the probability that a candidate pair of proteins form a complex.

We applied the P-HIPSTER algorithm to 12,237 viral proteins from 1,001 viruses spanning all viral families known to infect humans and 20,113 human proteins ([Figure S1A](#); [Table S1](#)). Of the 12,237 viral proteins interrogated, we compiled structural information for 7,593 (62%) represented by a total of 799 experimentally determined structures from the PDB and 109,888 homology models. We generated multiple models per viral protein to account for possible conformational heterogeneity and variable modes of protein binding. The structural coverage ranges

Table 1. Evaluation of the Overlap between High-Throughput Methods and P-HIPSTER

Dataset	Technique	Virus	PPIs/RNAi "Hits" Considered by P-HIPSTER	Predicted PPIs	Overlap
Shah et al., 2018	AP-MS	DENV2	190	181	4 (9×10^{-2})
Batra et al., 2018	AP-MS	EBOV	170	209	3 (0.26)
Calderwood et al., 2007	Y2H	EBV	111	956	12 (7×10^{-3})
Ramage et al., 2015	AP-MS	HCV	134	170	1 (0.68)
Germain et al., 2014	AP-MS	HCV	98	170	6 (1.9×10^{-4})
de Chassesey et al., 2008	Y2H + literature	HCV	481	170	21 (6.8×10^{-10})
Union of Ramage, Germain, de Chassesey	Multiple	HCV	606	170	25 (6.4×10^{-11})
Jäger et al., 2011	AP-MS	HIV	422	509	33 (1.3×10^{-8})
PPI-DBs	Multiple	Influenza A	326	132	11 (1.1×10^{-5})
Uetz et al., 2006	Y2H	KSHV	13	937	4 (2×10^{-2})
Davis et al., 2015	AP-MS	KSHV	555	937	40 (5×10^{-3})
Shah et al., 2018	AP-MS	ZIKV	189 ^a	89 ^a	0 (0.42)
Scaturro et al., 2018	AP-LC-MS	ZIKV	383	97	3 (0.28)
Ma et al., 2017	CRISPR/Cas9	EBV	142	956	14 (8.1×10^{-3})
Tripathi et al., 2015	RNAi	Influenza A	846	469	56 (2.4×10^{-12})
Jäger et al., 2011	RNAi	HIV	1,031	509	38 (9×10^{-3})

The number of viral-interacting human proteins or RNA "hits" correspond to those reported in the original publication and considered in the P-HIPSTER dataset. The set of influenza A PPIs were obtained from publicly available databases.

^aPPIs involving ZIKV capsid, NS3, or NS5 proteins.

from 73% in (+)single-stranded RNA (ssRNA) viruses to 33% in double-stranded DNA (dsDNA)-RT (retro-transcribing) viruses (Figures S1B and S1C) and represents a significant expansion on the limited knowledgebase for viral protein structures.

The PrePPI algorithm, on which P-HIPSTER is based, combines structural and non-structural clues to yield a final LR for PPIs within a given genome. For PrePPI, a physical interaction was found to occur with high confidence when the final LR ≥ 600 and the combined structural evidence yielded an LR ≥ 100 . Because P-HIPSTER is based entirely on structural evidence, we empirically tested the extent to which an LR cutoff of 100 is evidence for a true physical interaction. Sixty-five pairwise predictions spanning LR values from 1 to 1,106 were tested by co-immunoprecipitation (coIP). Of the 34 predictions with an LR >100 , 26 were confirmed to be positive (Figures 1B and S2; Table S2). Additionally, of the 31 predictions with LR <100 , only 7 were positive. These data correspond to a true positive rate of 79% and a true negative rate of 77%, similar to results obtained with the original PrePPI algorithm (Zhang et al., 2012). Furthermore, they provide a striking validation of the P-HIPSTER algorithm and indicate that an LR cutoff of 100, is indeed reasonable in identifying true direct interactions.

Of the 245,122,781 potential viral-human PPIs, P-HIPSTER reports LR values for 99,639,431 (Table S1). Of these, 282,528 (~0.1%) predictions have an LR ≥ 100 (data are available for detailed browsing at <http://phipster.org>). These are mediated primarily through interactions between structured domains (64.1%) and involve 7,463 viral proteins from 990 viruses and 5,749 human proteins. The majority (76.8%) of predicted viral-human PPIs mediated by domain-domain contacts are based on experimentally solved interaction complexes that do not

involve a viral protein. On average, 38 and 285 PPIs are predicted per viral protein and virus, respectively, with variability due to proteome size and structural coverage (Table S1). By comparison, structure-informed predictions from the PrePPI pipeline identify an average of ~6 interactions among proteins in the human proteome, in line with observations that viral proteins tend to mediate a greater number of interactions than human proteins (Garamszegi et al., 2013).

We compared P-HIPSTER predictions with results from high-throughput experimental approaches, yeast two-hybrid, and mass-spectrometry (Table 1). While the extent of overlap is modest, in most cases it is statistically significant and comparable to that found between different studies carried out on the same virus. As noted by others, limited overlap between independent experimental mapping of PPIs has been common and largely attributed to fundamental differences in experimental and statistical approaches that can influence rates of false-positives and false-negatives (Luck et al., 2017; Shah et al., 2018; von Mering et al., 2002). This, at least in part, may explain why P-HIPSTER results overlap well with some high-throughput studies and not with others.

We also compared P-HIPSTER predictions to results obtained from genome-wide loss-of-function screens used to identify cellular modulators of viral replication. The overlap between P-HIPSTER predictions and cellular factors for EBV (Ma et al., 2017), HIV (Jäger et al., 2011), and influenza (Tripathi et al., 2015) previously identified as "hits" in genome-wide screens were found to be highly significant ($p = 8.1 \times 10^{-3}$, 9×10^{-3} , and 2.4×10^{-12} , respectively; Table 1). Importantly, while overlap between independent RNAi screens has been relatively poor (largely attributed to disparities in cell types and assays used to

determine phenotypic consequences of knockdown), functional analysis on “hits” has demonstrated high concordance at an ontological and pathway level. Indeed, integrating virus-host PPIs together with pathway analysis of independent RNAi screens enables discovery of critical host factors that modulate viral replication (Tripathi et al., 2015). Because P-HIPSTER leverages structural homology to generate high-confidence predictions for viral-host PPIs, it captures interactions independently of cell type or assay used and can therefore be leveraged to explore and discover previously unappreciated circuitry that controls virus-human relationships.

Novel Mediators of Zika Virus Infection

Zika virus (ZIKV), an arthropod borne Flavivirus, has raised major health alarms due to devastating complications associated with infection, including Guillain-Barre syndrome and birth defects like microcephaly as well as spontaneous abortion and stillbirth (Musso and Gubler, 2016). Recent concerted experimental efforts have mapped ZIKV-human PPIs and identified cellular factors that mediate immune function and regulate viral replication (Grant et al., 2016; Li et al., 2016; Musso and Gubler, 2016; Scaturro et al., 2018; Shah et al., 2018; Tang et al., 2016; Wu et al., 2016). We leveraged P-HIPSTER predictions to augment these advances and identified 159 ZIKV-human PPIs, involving six viral and 97 human proteins, at an LR ≥ 100 (Figure 2). While the majority of these were not identified through previous experimental efforts, they are consistent with observations and functions associated with ZIKV infection (Grant et al., 2016; Li et al., 2016; Wu et al., 2016). Specifically, we find that the predicted ZIKV interactome is enriched for: (1) genes expressed in fetal brain cortex ($p = 1.16 \times 10^{-3}$); (2) genes known to be part of the transcriptional program in response to ZIKV infection in embryonic mouse brains ($p = 1.2 \times 10^{-3}$); and (3) immune-related pathways like antigen processing-cross presentation, interferon (IFN)- γ signaling, and nuclear factor κ B (NF- κ B) signaling (Figure 2B). We experimentally confirmed 12 (of 16 tested interactions) novel ZIKV-human interactions and modeled the interaction mode for 7 of them (Figure 2C). The prediction and validation of interactions not previously observed through traditional experimental efforts underscores the sparsity of relationships reported in the literature, the susceptibility of such approaches to incomplete mapping of PPIs, and the value of combining experimental techniques (like MS and Y2H) with computational predictions provided by P-HIPSTER.

We combined predicted ZIKV-human PPIs along with known human PPIs from IntAct into a single protein interaction network and identified 9 topological modules: four modules composed of proteins predicted to interact with >1 viral protein, and 5 modules defined by cellular proteins with only a single ZIKV protein partner (Figure 2B; Table S3). Modules predicted to interact with >1 viral protein are enriched for functions associated with apoptosis, embryonic lethality, cell-cycle, abnormal neuron proliferation, and conjunctivitis—consistent with known biological pathways and phenotypes associated with ZIKV infection (Li et al., 2016; Musso and Gubler, 2016; Scaturro et al., 2018; Shah et al., 2018; Tang et al., 2016). In addition to recapitulating known biology, we find that the NS3 sub-network is enriched for regulators of peptidase activity—suggesting that they play

a critical role in NS3 function. Modules specific to ZIKV envelope (Env) or capsid (C) are associated with immune-related pathways. While the data implicate capsid in modulating Toll-like receptor and RIG-I mediated signaling, two important modes of pattern recognition that play indispensable roles in innate immunity and initiate cellular responses to ZIKV infection (Hamel et al., 2015), Env function is associated with T cell responses (Figure 2B)—ones that are critical for control of, and long term protection from, ZIKV infection (Lima et al., 2017).

We selected six cellular factors with validated ZIKV-human PPIs (Figure 3; Table S2) for experimental interrogation through gain and loss-of-function (Figure 3). We observe that while overexpression of human amyloid β precursor APP results in a modest (yet significant) 2-fold enhancement of ZIKV replication, it dramatically affects cellular response to IFN β stimulation (Figure 3A). While priming of cells with IFN β normally induces a refractory state (resulting in a 10-fold reduction in viral titers), priming with IFN β has no effect on cells exogenously expressing APP. APP is a known partner of Musashi-1 (MSI1), a cellular protein recently shown to regulate ZIKV replication through interaction with the viral genome and repression of genes involved in neural stem cell function (Chavali et al., 2017). This suggests a potential role for the experimentally confirmed PPIs between APP and ZIKV C and NS3 in modulating MSI1-mediated control of viral replication and cellular responsiveness to IFN β .

In contrast to APP, we find that while SFN (14-3-3 σ) overexpression results in ~ 10 -fold increase in viral titers, it potentiates cells to IFN β priming and enhances the cellular antiviral state (Figure 3A). 14-3-3 ϵ , which is related to SFN, has recently been shown to regulate RIG-I localization (Liu et al., 2012) and to be a target of dengue virus NS3 protein—an interaction that antagonizes RIG-I signaling and results in enhanced viral replication (Chan and Gack, 2016). Similarly, SFN and 14-3-3 ϵ have also been shown to inhibit Toll-like receptor (TLR)-mediated sensing of viral RNA (Butt et al., 2014), and 14-3-3 ζ was demonstrated to interact with STAT3 and promote downstream signaling (Han et al., 2015). Our findings suggest a broader role for this protein family in controlling innate immune responses to viral infection and implicate SFN in regulating both cell-intrinsic antiviral programs, as well as those induced by IFN β .

The most dramatic impact on viral replication resulted from overexpression of ESR1, which led to $\sim 2,000$ -fold reduction in ZIKV replication (Figure 3A). Conversely, small interfering RNA (siRNA) knockdown of ESR1 (that resulted in 70% depletion of ESR1 mRNA) potentiated ZIKV replication as measured by both viral titer and viral mRNA in infected cells (Figures 3B and 3C). Although the precise molecular process through which ESR1 regulates ZIKV replication is yet to be defined, the discovery that FDA-approved selective estrogen receptor modulators to inhibit Ebola virus replication (Johansen et al., 2013) suggests a broader role for this hormone in modulating viral replicative potentials. Moreover, recent epidemiological data suggest different ZIKV incidence rates between men and women as well as pregnant versus non-pregnant women (Lozier et al., 2016). While potential explanations include exposure to *Aedes* mosquitoes, increased severity of symptoms in women versus men, reporting biases, and sexual transmission, our findings

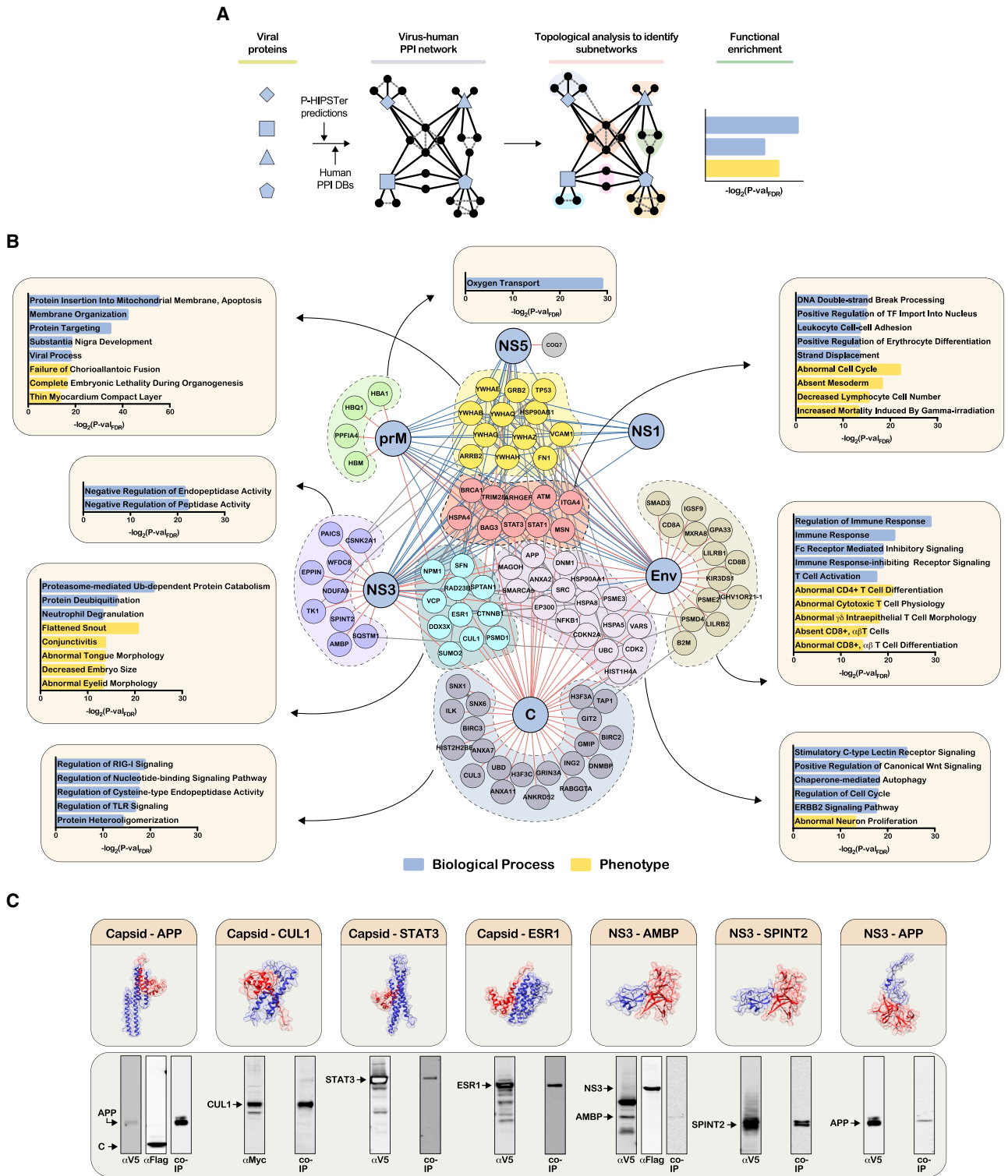


Figure 2. Discovery of Novel ZIKV-Human Protein Interactions

(A) Combining P-HIPSTER predictions with known human PPIs followed by topological analysis identifies connectivity-based modules that are subjected to functional interrogation.

(B) ZIKV-human PPI network with enriched biological pathways and phenotypes for each topological module.

(C) Interaction models and experimental validation for 7 predicted ZIKV-human PPIs (blue and red models correspond to human and viral proteins, respectively). See also [Figure S2](#) and [Table S3](#).

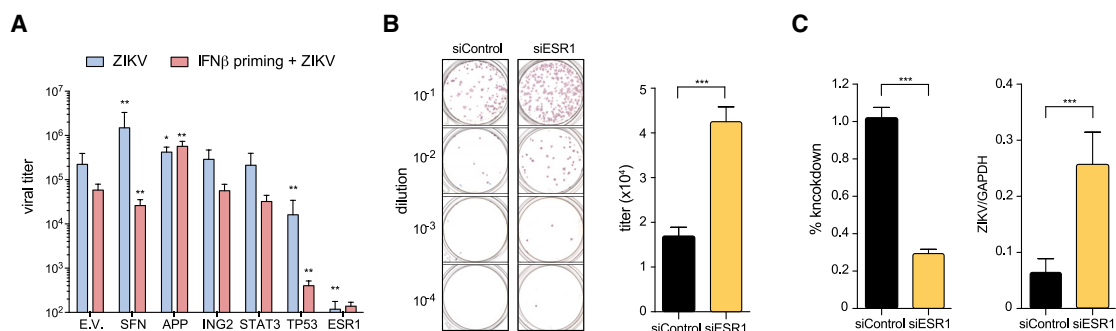


Figure 3. Functional Interrogation of ZIKV Cellular Partners and Identification of ESR1 as an Inhibitor of Viral Replication

(A) Effect of overexpressing cellular factors on ZIKV infection in 293T cells with or without IFN β priming. Stars indicate significant difference in viral titer (determined by focus forming assay) or cellular response to IFN β .

(B and C) Effect of ESR1 knockdown (siRNA) on ZIKV replication. (B) Focus forming assay (red staining indicates ZIKV foci). (C) qPCR of ZIKV mRNA, data representative of 3 independent experiments. * $p < 0.05$; ** $p < 0.0001$. EV, empty vector.

See also [Figure S2](#) and [Tables S2](#) and [S3](#).

reveal an additional point of viral vulnerability that may be leveraged for therapeutic targeting of the estrogen pathway. In short, functional interrogation of experimentally validated ZIKV-interacting human proteins reveals novel roles for three human proteins (APP, SFN, and ESR1) in modulating viral replication and/or cellular responsiveness to IFN β . Thus, P-HIPSTER predictions for a virus of interest can be leveraged to uncover novel PPIs, identify interactions underlying clinical phenotypes of viral infection, expose regulators that control inflammatory responses and viral replication, as well as identify targets with wide therapeutic potential.

Classifying HPVs Based on PPIs

HPVs, non-enveloped dsDNA viruses that preferentially infect epithelial tissues, can cause lesions with various degrees of severity. HPVs encode two sets of proteins defined by their temporal expression during the viral replicative cycle: (1) late proteins (L1–L2) form the icosahedral capsid; and (2) early proteins (E1–E7) serve regulatory functions including replication control (E1 and E2), cell-cycle regulation, immune evasion, and virus release (E4–E7). E6 and E7 are transcriptionally regulated by E2 and have been implicated in dictating viral oncogenicity (Doorbar et al., 2012). Low-risk (LR) HPVs have been linked primarily with benign warts while high-risk (HR) HPVs can lead to neoplasias and carcinomas like cervical cancer (Cubie, 2013). HR and LR oncogenic classifications have largely relied on epidemiologic and phylogenetic information which is essential for prevention planning and screening programs (Muñoz et al., 2003). While vaccines against HPV exist, they have no therapeutic value against existing infections and must therefore be administered before individuals are infected. So, identification of viable therapeutic targets is critical and delineating cellular proteins that are differentially targeted by HR and LR HPVs can be particularly useful in identifying circuits that mediate disease.

We applied a supervised feature selection method, using PPIs with LR ≥ 100 as input, to five LR-HPVs and five HR-HPVs associated with cervical cancer with the goal of identifying PPIs that discriminate HR and LR infections (Figure 4; Table S4). Evalua-

tion of the Bayesian classifier by 5-fold and leave-one-out cross-validation yielded an accuracy of 90% in predicting cervical cancer risk for a given HPV and resolved ten human proteins whose interactions differentiate LR- and HR-HPVs (Figure 4B). HR-HPVs E7 and E2 proteins have a propensity to interact with CCNB1, SMAD3, PCNA, HNRNPM, and HNRNPDL (group I) while LR-HPVs E1 and E2 proteins preferentially interact with LIRB1, TERF1, FUBP1, BRDT, and TK1 (group II). In both cases, LR-values clearly distinguish HR- and LR-HPVs.

We also identified 18 proteins that interact with both HR-HPVs and LR-HPVs (group III), 10 of which have been shown to have direct roles in the HPV life cycle (Buitrago-Pérez et al., 2009; Collier et al., 1998; Dietrich-Goetz et al., 1997; Grinstein et al., 2002; Kajitani and Schwartz, 2015; Katzenellenbogen et al., 2010). Among these, P-HIPSTER recapitulates previously reported interactions between HPVs and RB1, RBL1, RBL2, and p53 (Buitrago-Pérez et al., 2009). Indeed, p53 and RB family members play well-accepted roles in HPV oncogenesis. The results discussed below suggest that the interaction of viral proteins with group I and II proteins may modulate the established roles of group III proteins in determining HPV oncogenic potentials.

Multiple lines of evidence connect group I proteins, including CCNB1, SMAD3, and PCNA, with an augmented risk of cervical intraepithelial neoplasia (CIN) and cervical cancer (Cho et al., 2006; Lee et al., 2002; Tjalma et al., 2001). CCNB1, predicted to preferentially interact with E7 protein of HR-HPVs (Table S4), plays an important pro-mitotic role by promoting the G2/M transition (Takizawa and Morgan, 2000)—elevated levels have been demonstrated in multiple cancers, including cervical cancer (Cho et al., 2006). While previous reports illustrate that E7 promotes mitosis and cellular proliferation through interactions with cellular B-Myb-MuvB (Pang et al., 2014), our results suggest that, in addition, E7-CCNB1 interaction may contribute to fulfilling its pro-mitotic role. In agreement with previous reports, we also identify SMAD3, a signal transducer and transcriptional modulator of transforming growth factor β (TGF- β) signaling that promotes tumor suppression, as a partner of E7. HR-HPV16 E7 has been shown to interfere with TGF- β induced cell

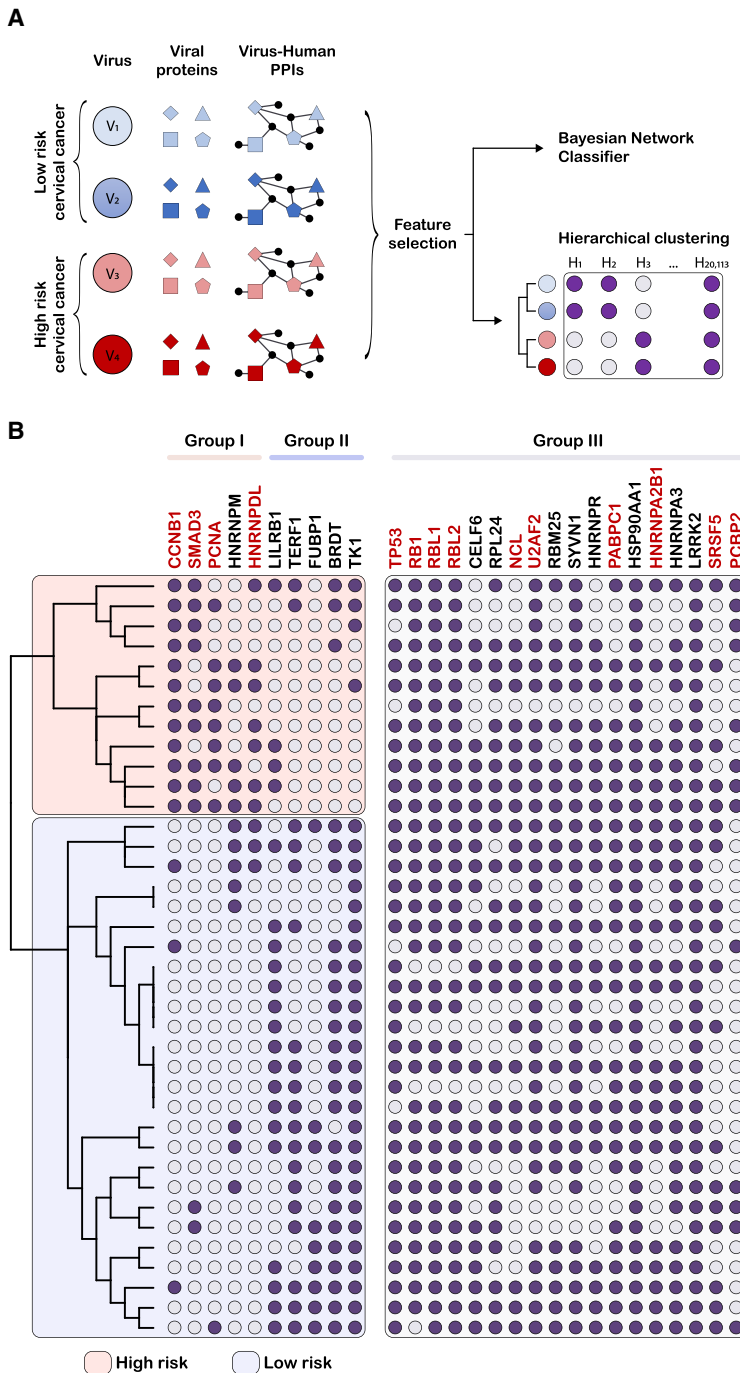


Figure 4. P-HIPSTER-Derived Bayesian Network Classifier Discriminates High- and Low-Risk HPVs

(A) Machine learning on P-HIPSTER interactomes for HR and LR HPVs (highlighted in B with red and blue rectangles, respectively) is used to identify features associated with viral oncogenic potential.

(B) Hierarchical clustering of alpha HPVs based on a constellation of 10 viral-host PPIs (5 associated with HR, 5 associated with LR; group I and group II, respectively) discriminates HR and LR HPVs. Group III describes 18 viral-host PPIs shared across alpha HPVs. Dark and open circles denote binding profiles (LR ≥ 100, dark circle; LR < 100, open circle). Human proteins with known roles during HPV infection are highlighted in red.

See also Table S4.

In addition to CCNB1 and SMAD3, we identified PCNA as a cellular partner of HR-HPV E2 protein but not LR-HPVs (Table S4). Expression of PCNA has been associated with HR-HPV and is correlated with increasing grade of CIN and cervical cancer (Branca et al., 2007; Eissenberg et al., 1997; Kelman, 1997; Tjalma et al., 2001). The large difference in LRs observed for this interaction resides in the peptide-domain interaction, where E2 sequences from HR-HPVs match the peptide motif of an ELM class while the E2 LR-HPVs do not (Table S4). During HPV infection, viral proteins recruit DNA polymerases and members of the DNA replication machinery, including PCNA, to trigger viral replication (Berg and Stenlund, 1997; Chojnacki and Melendy, 2018; Fuss and Linn, 2002; Melendy et al., 1995; Mohr et al., 1990). Indeed, HPV is known to indirectly regulate the activity of PCNA through an interaction between viral E7 and human p21 protein (Funk et al., 1997). While the precise mode of action remains to be elucidated, our results suggest that the E2-PCNA interaction contributes to HPV-mediated modulation of genome replication and initiation of oncogenic states.

The remaining 2 human proteins preferentially targeted by HR-HPVs are HNRNPM and HNRNPDL (via viral E2 protein). In addition to acting as a transcriptional repressor of viral E6 and E7 proteins, E2 contributes to RNA processing and metabolism (Kajitani and Schwartz, 2015). Notably, HNRNPM and HNRNPDL belong to the hnRNP family that participates in various aspects of RNA metabolism including alternative splicing, mRNA stabilization, and regulation of transcription and translation (Geuens et al., 2016), and their expression is associated with various cancer types (including cervical carcinoma for HNRNPDL) (Chen et al., 2014; Sun

growth by binding SMAD3 and blocking its transcriptional activity (Lee et al., 2002), our results extend this regulatory potential to E7 proteins of other HR-HPVs. Indeed, recent reports suggest that E7 may modulate oncogenic potential of HPV's (Mirabello et al., 2017; White and Munger, 2017). Thus, P-HIPSTER provides a basis for further understanding the role of E7 in carcinogenesis and points to additional layers of regulation through which HPVs modulate viral replication and initiation of oncogenic states in infected tissues.

et al., 2017; Tsuchiya et al., 1998). While the role for HNRNPM, a component of the spliceosome complex, remains to be determined, HNRPNL has been implicated in inhibiting the production of HPV-16-spliced L1 (HPV major capsid protein) mRNA and promoting immune system evasion and establishment of long-term persistent infections with enhanced risk of carcinogenesis (Li et al., 2013).

Group II proteins point to unappreciated biological processes underlying HPV pathogenesis. While none are known to play a direct role in the HPV life cycle, changes in TK1 and FUBP1 expression have been associated with cervical cancer (Buitrago-Pérez et al., 2009; Chen et al., 2013; Pyeon et al., 2007). The relative sparsity in prior knowledge likely reflects the fact that LR-HPVs are far less studied than HR-HPVs. Our results invite new studies into group I and II proteins in cancer progression and underline the importance of evaluating host factors in the context of both LR- and HR-HPV.

Having established the Bayesian classifier on five known HR-HPVs and five known LR-HPVs, we clustered the remaining 28 HPVs based on predicted PPIs with the 10 human proteins in groups I and II. The resulting dendrogram discretizes HR-HPVs and LR-HPVs into two branches that are supported by known biology (Figure 4B). Of the seven HPVs that co-cluster with HR-HPVs, four have been described as potentially carcinogenic (HPV-53, HPV-69, HPV-82, HPV-97) (Cubie, 2013; IARC, 2019), two (HPV-68a and HPV-68b) are subtypes of high risk HPV-68, and one (HPV-62), although considered to be a LR-HPV, is commonly found in neoplastic tissue, is detected in coinfection with HR-HPVs, and is among the most prevalent LR-HPVs found in women with abnormal cervical cytology or cervical cancer (Artaza-Irigaray et al., 2017). Of the 21 HPVs that cluster with LR-HPVs, only one has been described as potentially carcinogenic (HPV-67) (IARC, 2019).

The classifier presented here is orthogonal to the standard one based on sequence alone and can be used in parallel when new viruses are discovered. However, in addition to providing a demonstration of P-HIPSTer performance, the results identify previously unappreciated PPIs that discriminate between HR- and LR-HPVs. To our knowledge, this has not been accomplished previously and offers a new perspective on HPV pathogenesis, oncogenic potential, and cellular factors that may serve as viable therapeutic targets.

Cellular Pathways Coopted by Viruses

Beyond utilizing P-HIPSTer to gain insights into specific viruses of interest, its comprehensive nature also offers a unique opportunity to illuminate a broad picture of viral infection and identify features that are shared across human viruses. For example, Gene Ontology (GO) analysis across the 5,749 cellular factors predicted as viral-interacting proteins indicates significant overrepresentation of biological processes related to signal transduction, immune response, and viral infection (Figures S3A and S3B). We also identified 173 human proteins, predicted to interact with at least 100 viruses, enriched for processes related to immunity, virus infection, and functions known to be important during viral replication (e.g., regulation of cell cycle and membrane organization), providing insights into cellular processes broadly relevant for virus-human dynamics (Figures S3C and

S3D). In agreement with recent findings (Chen and Xia, 2019), we find that these proteins are also enriched for virally implicated genetic diseases. The identification of pan-viral factors that converge around common cellular pathways is made possible both by the comprehensive nature of the P-HIPSTer database and by the large number of interactions detected by structural homology.

In addition to the standard analysis used to identify overrepresented biological themes and ontologically related gene groups, we previously demonstrated that gene set enrichment analysis (GSEA) (Subramanian et al., 2005) can be used to provide functional annotation of a given protein based on the functions of its predicted binding partners (Garzón et al., 2016). We applied this analysis to the LR-based rank-order of all predictions for a given viral protein and represented the pathways and functions targeted by a given virus as the union of enrichments across its proteins (Figure S4). This exposed 190 pathways that are recurrently targeted by DNA, RNA, and RT viruses, 92 of which are associated with regulation of cellular metabolism and point to specific metabolic processes underpinning viral requirements during infection (Table S5). For example, glucose metabolism and fatty acid synthesis are universally targeted across viruses, reflecting the requirement for increased membrane production necessary for viral packaging, vesicular transport, and protein production and the need for energy through glycolysis and beta-oxidation of fatty acids (González Plaza et al., 2016). We also find that the majority glycan-related metabolic pathways are targeted by viruses across all nucleic acid types (Table S5), reflecting their central roles in immune evasion (through modulation of adaptive immune responses to infected cells) and supporting pathogen recognition by antibodies (Raman et al., 2016).

Underscoring the universal role that Jak-Stat, NF- κ B, and type-I IFN-dependent pathways play in initiating innate immune responses to infection, we find that they too are recurrently targeted across all viruses (Figure S4B; Table S5). This is further emphasized by the observation that among the 173 human proteins predicted to interact with ≥ 100 viruses, 55 contain a sushi domain—known to be involved in the complement cascade, innate immune cell trafficking and viral entry, and endocytosis (Figure S3D) (Ley, 2003; Tanner et al., 1988). Moreover, targeting of complement pathway and complement and coagulation cascades is enriched among RNA and DNA viruses. Yet, while viruses converge around response-initiating pathways and share metabolic requirements for infection, replication, and pathogenesis, they have diversified in their targeting of other arms of the immune system (only 16 out of 182 immune related pathways are commonly targeted by all viruses). For example, poxviruses widely target immune-related pathways, while noroviruses specialize in targeting innate immune and type I IFN responses but do not widely interact with proteins involved in nucleic acid sensing. Conversely, our results indicate that members of other viral families have diversified in their targeting immune system components. For example, flaviviruses utilize one of two distinct strategies in targeting immune related pathways; while the majority of flaviviruses (including dengue, Zika and West Nile viruses) interact with pathways involved in T cell differentiation, DAP12-mediated natural killer (NK) cell responses, and RIG-I/MDA5 signaling, hepaciviruses target eicosanoid ligand binding

receptors and IL-23 signaling (Figure S4B). These observations invite questions about the selective pressures that gave rise to such diversification in adaptive strategies within a family of related viruses. Furthermore, these data highlight the power of P-HIPSTER to capture general and specialized strategies employed by viruses to rewire cellular responses to infection and to uncover principles that govern human-virus relationships that were otherwise hidden or missing from other PPI discovery methods.

Reconstructing Functional and Evolutionary Relationships across Human-Infecting Viruses

In addition to providing insight into shared and unique strategies exploited by human viruses, knowledge of PPIs informs about functional and evolutionary relationships that cannot be discerned through sequence alone. For instance, sequence and structural comparison of viral proteins illustrates that while sequence homology occurs between proteins from viruses of the same nucleic acid type, high structural similarity is often found for protein pairs belonging to viruses of different nucleic acid types (Figure S5). We employed an unsupervised clustering strategy (where viruses are described by their set of enriched biological pathways, given their corresponding set of predicted viral-human PPIs) to uncover functional relationships that are masked by sequence divergence (Figure 5A). We identified eight distinct clusters, each representing a unique constellation of pathways targeted by member viruses (Figure 5; Table S6)—revealing novel insights into shared and unique infection strategies employed across virus families.

Three clusters are dominated by viruses belonging to one of three families (cluster 1: *Picornaviridae*, cluster 5: *Caliciviridae*, cluster 8: *Poxviridae*), suggesting specialized infection strategies that distinguish them from other human viruses. Poxviruses are characterized by pathways related to signaling, immunity, transcription, and cell growth. In particular, we observe that *Poxviridae* family members target interleukin-10 (IL-10), an anti-inflammatory cytokine that regulates immune responses to mitigate tissue damage and is known to be modulated by viruses to evade immunity (Ouyang et al., 2014) (Figure 5B; Table S6). While several poxviruses have been shown to encode an IL-10 ortholog (Ouyang et al., 2014), we find that this family of viruses recursively target the IL-10 pathway at multiple levels of regulation including production and downstream signaling, reflecting a particular importance of this pathway during infection.

P-HIPSTER can also help contextualize findings from mouse model systems and when considering experimentally intractable viruses or viruses for which not much more than genome sequence is known (Karst et al., 2014). Noroviruses (*Caliciviridae* family) have emerged as human pathogens of significant public health import. These viruses do not replicate well *in vitro* and much of what is known about them comes from studies of related mouse caliciviruses, resulting in limited understanding of cell factors that are involved in human infection (Wobus et al., 2006). As (+)ssRNA viruses, noroviruses regulate mRNA translation machinery to facilitate viral replication and modulate translation of certain host mRNAs, including interferon stimulated genes (ISG) (Emmott et al., 2017). We find that in addition to pathways related to immunity and regulation of the cellular extracellular matrix, all biological

pathways related to mRNA processing are enriched within noroviruses (cluster 5) (Figure 5B; Table S6), suggesting a pervasive and systematic cooption of cellular machinery involved in RNA metabolism by this family of viruses.

Clusters 2, 4, 6, and 7 bring together viruses from different families and Baltimore categories. Strategies employed by viruses within these clusters appear to have converged around biological pathways of shared importance during infection (by interacting either with the same host proteins or with host proteins of the same pathway). A similar observation, albeit at smaller scale, was described when clustering viral proteins based on their proteomic profiling and network proximities of interacting partners (Pichlmair et al., 2012). Cluster 7, which contains all Ebola and measles viruses, and evolutionarily distant viruses, converges on a variety of pathways related to signaling, immunity, apoptosis, and cell growth (Table S6). Of these, Rho GTPase signaling ($p = 5.1 \times 10^{-36}$), regulates actin and microtubule dynamics and biological processes like phagocytosis, cellular transport, and intracellular communication, by linking membrane receptors to the actin cytoskeleton (Van den Broeke et al., 2014) (Figure 5). While Rho GTPases facilitate Ebola virus entry through macropinocytosis, measles virus interferes with this pathway to perturb T cell function and long-term T cell-mediated immunity (Müller et al., 2006; Quinn et al., 2009). Our observations extend a role of Rho GTPase signaling to other members of this cluster and illustrate that although some viruses converge around particular cellular pathways, utilization of such machinery is dependent on cellular context and viral requirements for their replicative cycle.

We also find functional divergence among some viral families. For example, flaviviruses appear in multiple clusters, including clusters 2, 4, and 6. Viruses in cluster 2 converge around pathways related to transcriptional and translational regulation (e.g., “transport of mature transcript to cytoplasm” and “mRNA capping”). While Zika virus falls into cluster 2, phylogenetically related dengue viruses fall into cluster 4, where signaling and immune-related pathways are among the most enriched pathways (e.g., “immune response IFN α/β signaling”). Instead, viruses in cluster 6 (e.g., Japanese encephalitis virus) are enriched for a different set of immune pathways, cell growth, and apoptosis-related pathways. In short, clustering viruses based on pathway enrichments derived from P-HIPSTER inferred PPIs enables discovery of convergent functional relationships between unrelated viruses and divergent relationships within viral families. Moreover, clusters that bring together disparate viruses demonstrate that while sequence divergence precludes any possibility of using traditional methods to go beyond ontological classification, structure-based PPIs afforded by P-HIPSTER add a level of cross-stratification to virus taxonomy by identifying functional relationships that are otherwise masked. Finally, the data open the possibility of discovering molecular processes that underlie pathological states and clinical outcomes across unrelated viral infections.

History of Selection Imposed by Viruses

In the context of viral infection, as with other pathogens, there exists an “arms race” between the virus and infected cells—signs of which can be found in genomes of both virus and host. We interrogated genetic divergence among primates as a proxy for the selection pressures imposed by viruses on the

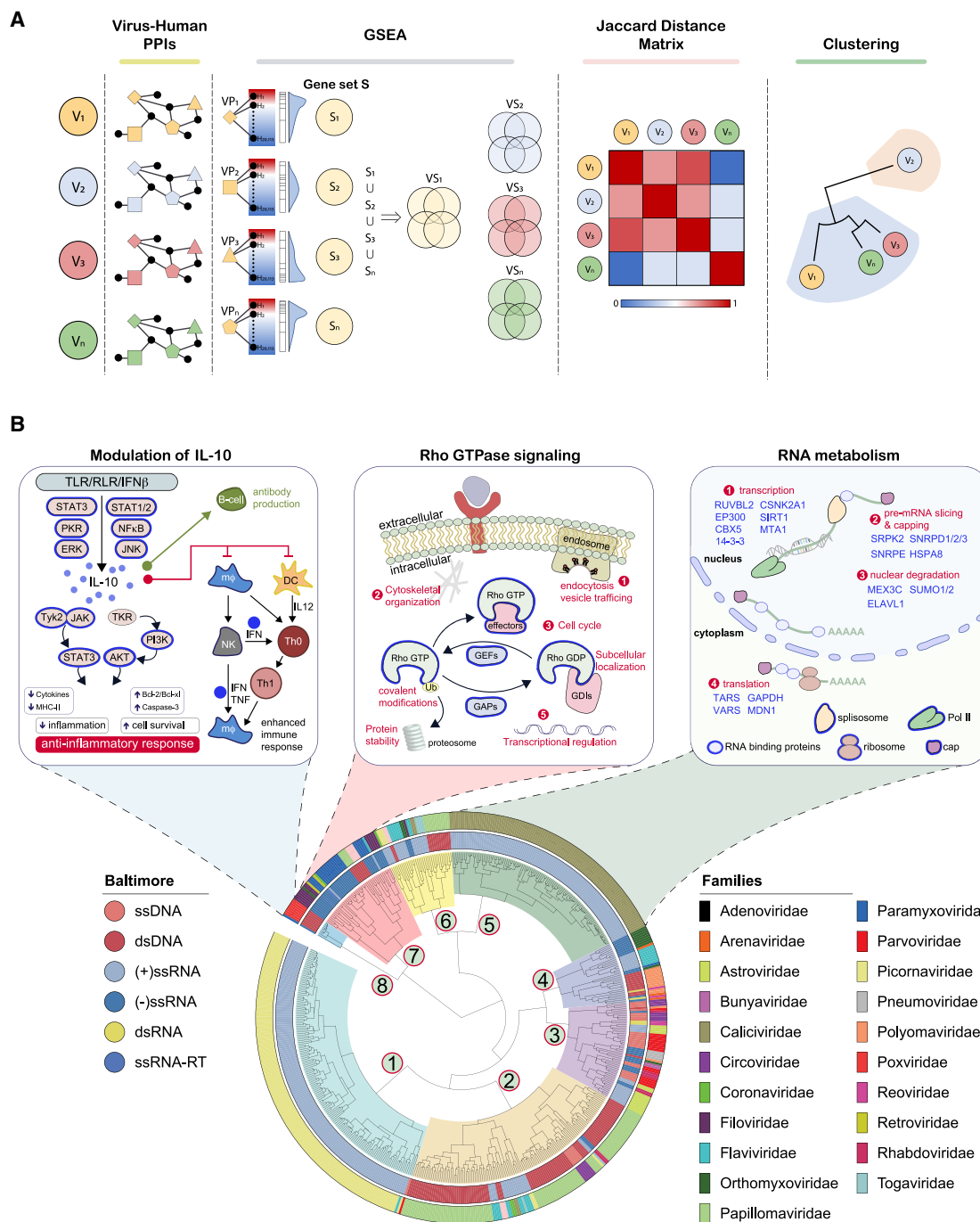


Figure 5. P-HIPSTER Reveals Functional and Evolutionary Relationships across the Human Virome

(A) Strategy to cluster viruses, described by their set of enriched pathways.

(B) Dendrogram of 568 viruses clustered based on their enriched pathways. Inner color ring specifies Baltimore category; outer ring specifies viral family for each virus. Schematic representations of the interleukin 10 (IL-10; enriched in Cluster 8), Rho GTPase signaling (enriched in cluster 7), and RNA metabolism (enriched in cluster 5) pathways. Pathway components predicted to be targeted by viruses within each cluster are highlighted in blue.

See also [Figures S3, S4, and S5](#) and [Tables S5 and S6](#).

human proteome (Figure 6). As noted by many others (Bustamante et al., 2005; Daugherty and Malik, 2012; Hughes et al., 2003), the primate lineage is dominated by purifying selection.

We found that proteins predicted to interact with viral proteins display significantly lower $\Delta N/\Delta S$ values (average of 0.23 versus 0.31; Wilcoxon tests: $p < 0.0001$, $W > 10^6$) than non-interacting

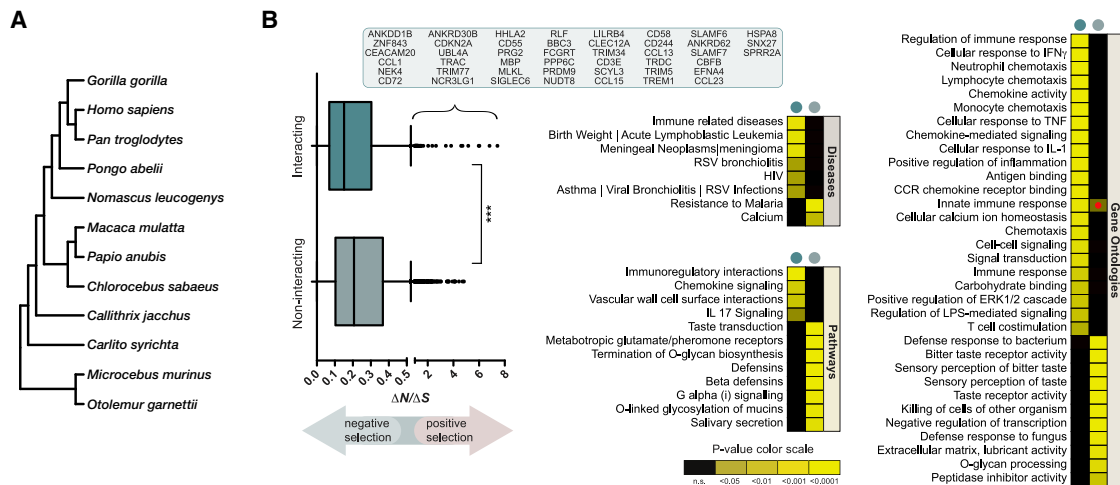


Figure 6. P-HIPSTER Reveals History of Selection Imposed by Human Viruses

(A) Evolutionary tree derived from 12 whole genome primate sequences.

(B) $\Delta N/\Delta S$ values for each of 14,974 aligned genes were plotted across human-virus interacting and non-interacting proteins. Listed in the inset are virus-interacting proteins with $\Delta N/\Delta S$ values >1 . Shown are functional enrichments ($p < 0.05$ and $FDR < 20\%$) for interacting (dark circle) and non-interacting (light circle) proteins with $\Delta N/\Delta S > 1$. Red dot denotes p value that did not pass FDR correction. Stars indicate Wilcoxon $p < 0.0001$, $W > 10^6$.

See also Table S7.

proteins. The observation suggests that virus-interacting proteins are under stronger purifying selection than non-interacting ones. This observation is in agreement with previous results using orthologous, albeit smaller, datasets (Enard et al., 2016; Halehalli and Nagarajaram, 2015). We also find that proteins with the lowest $\Delta N/\Delta S$ values (5th percentile) that are predicted to interact with viruses are enriched for regulators of gene expression, chromatin modification, and cell cycle, in addition to modulators of innate and adaptive immune responses (Table S7). Although such increased conservation may reflect an over-representation of essential genes targeted by viruses, it points to critical proteins at the interface of viral-human interactions that span millions of years of primate evolution.

Positively selected genes are of particular interest, as they represent key adaptive sweeps that have been shaped by environmental pressures, including those imposed by viruses. Analysis of the 45 virus-interacting proteins under positive selection (with mean $\Delta N/\Delta S$ values >1) reveals that they are highly enriched for functions and diseases related to viral infection, as well as adaptive and innate immune responses (Figure 6B)—a similar relationship was recently described (Chen and Xia, 2019). Reflecting their central role in initiating and mediating responses to a variety of pathogens, regulators of immunity are also enriched among the 329 positively selected genes not predicted to interact with viruses. However, these genes are over-represented by functions and diseases related to bacterial, fungal, and parasitic infection, demonstrating that P-HIPSTER can distinguish among proteins that do and do not interact with viral proteomes and have functions associated with viral infection. Notably, TRIM5, previously identified as a host restriction factor for HIV that determines human versus macaque susceptibility (Sawyer et al., 2005), is among the 45 positively selected genes targeted by viruses. As noted by others (Barreiro and Quintana-Murci, 2010), genome-wide scans for molecular

“footprints” of natural selection imposed by infections would benefit greatly from integrative approaches that combine genetics, epidemiology, immunological phenotyping, and knowledge of key pathogen-host molecular interactions. Therefore, while the notion that human genetics has been shaped by pathogens is not new, P-HIPSTER data extend the list of human proteins shaped by their interactions with viruses and highlight its ability to capture evolutionary histories that have determined pathogen and disease susceptibility. Thus, adding to the toolbox of resources used to investigate evolutionary pressures imposed on the human genome and to interrogate evolutionary relationships between humans and their virome.

DISCUSSION

As viruses exploit host cell through molecular interactions involving diverse molecules including nucleic acids, lipids, sugars, and proteins, their study has resulted in the discovery of many fundamental aspects of cell biology. The work presented here represents the largest initiative to model and map cross-species protein-protein interactome, and as such, should serve as a valuable resource for studying viral replication strategies, evolutionary and functional relationships, and history of human adaptation. Furthermore, this compendium may accelerate the discovery of drugs with potential pan-viral effects by targeting host pathways coopted across multiple viruses. By providing structural models for the majority of viral proteins included in the study, and models for interactions mediated by them, P-HIPSTER invites in-depth exploration of interfacial residues, examination of the structure space utilized by viruses, and an analysis of the evolutionary constraints imposed on viral protein functions.

To facilitate convenient access to P-HIPSTER derived interactions, we have made the results available through an interactive webserver that enables both searchable queries and data

download (<http://phipster.org>). Specifically, the database includes: all interactions with $LR \geq 100$ (representing 282,528 PPIs between 12,237 viral and 20,113 human proteins), structure models for all modeled viral proteins, structure models for predicted complexes with a domain-based $LR \geq 100$, and navigational links to external databases. In addition, annotations for all viral proteins are provided based on known functions of structure neighbors and sequence homologs as well as functional analysis of P-HIPSTER-predicted interaction partners. In total, the repository provides over 100,000 atomic models corresponding to ~7,600 viral proteins from 970 viruses. Finally, as the structural information in the PDB continues to grow and provides critical sources of evidence for PPI predictions, the P-HIPSTER database will undergo annual updates. Because it relies on homology-modeling based on experimentally determined crystal structures, as the number and diversity of solved structures increases, so will P-HIPSTER's predictive power and scope. While some human infecting viruses and viral strains may be missing from the current compendium, PPI discovery and structural analysis for individual viruses of interest are possible upon request.

P-HIPSTER, like PrePPI and other high-throughput computational and experimental methods, is ultimately intended to generate testable hypotheses. The reliability of predictions contained in the database can be assessed in a number of ways. Along with *in silico* recapitulation of results obtained by orthogonal methods, and true positive and true negative rates above 75%, the applications discussed above demonstrate that P-HIPSTER yields predictions that either rediscover known biology or point to PPIs that are consistent with the expected effects of viral infection. Moreover, even in cases of well-studied viruses like ZIKV that have been subject to repeated experimental efforts to comprehensively identify viral-human PPIs, P-HIPSTER led to discovery and validation of novel and functionally relevant interactions, including ESR1 that we demonstrate to have a critical role in modulating ZIKV replication.

While P-HIPSTER rediscovers multiple previously identified interactions between viral and cellular proteins, others are not reported with an $LR \geq 100$. For example, HPV PPIs including E7-URB4 and E6-MAML1 were not identified. However, there is no evidence to suggest that these interactions are direct. Indeed, methods like affinity purification and chromatography, coIP, and fluorescence microscopy, used to identify these interactions and others, often do not discriminate between direct and indirect interactions. Because P-HIPSTER identifies direct interactions, these examples highlight the orthogonal and complimentary nature of the approach in mapping the extended network of cellular partners and complexes of viral proteins. Similarly, some false-positive P-HIPSTER predictions, notably HPV E7-p53 (known experimentally not to occur), will inevitably arise. Nevertheless, the high validation rate along with the discoveries highlighted in each of our applications of the pipeline engender confidence in P-HIPSTER reliability.

Although P-HIPSTER is orthogonal to traditional experimental methods of identifying PPIs, it offers several key benefits. First, while high-throughput methods can be influenced by experimental settings and generally fail to provide information about the nature of an interaction, P-HIPSTER provides context independent evidence of a direct physical interaction. Moreover, access to structure models for all interactions with a domain-

domain $LR \geq 100$ (85,175 in total) offers the ability, not available by other methods, to experimentally probe each prediction with site-directed mutants. Moreover, such models are extremely useful in generating functional hypotheses and interpreting genetic data on susceptibility to viral infection. While the current sparsity of data on human genetic diversity does not provide the statistical power to analyze differential selection pressure on interfacial versus non-interfacial residues, the proliferation of genome sequencing projects may soon make such an inquiry tractable. Second, P-HIPSTER exploits the fact that protein structure is far better conserved than sequence and provides new functional annotations for ~7,600 viral proteins—representing a massive expansion on incomplete annotations provided by current computational tools that rely on sequence identity. Most importantly, the new findings described in this work demonstrate the facility with which P-HIPSTER's comprehensive nature enables fundamentally new biological discoveries and insights that could not have been realized with existing resources.

P-HIPSTER's scalability is an important step toward the goal of defining all pathogen-host interactions across any and all species. Indeed, future developments will include exploration of endogenous retroviruses, eukaryotic and bacterial pathogens of humans and agricultural crops and livestock, as well as bacterial viruses that are part of the human enteric microbiome. In addition, efforts to leverage whole-genome sequencing together with P-HIPSTER predictions may systematically map the landscape of human genetic variation with endemic viral infection across geographic regions and inform on adaptive costs incurred during human evolution. Integration of P-HIPSTER data with other large-scale experimental mapping of pathogen-host interactions will augment and accelerate discovery of basic cell-biological machinery and targeted interrogation of both experimentally tractable and intractable viruses. Finally, affordable gene synthesis pipelines, coupled with tunable expression systems and low input profiling of cellular transcriptional states, may be used to bridge PPIs and reconstruct signaling pathways modulated by viral proteins of interest.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCE TABLE**
- **LEAD CONTACT AND MATERIALS AVAILABILITY**
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
 - Cell lines, virus strains
- **METHOD DETAILS**
 - Assembly of viral protein and human protein dataset
 - Protein modeling, structural neighbor search
 - Disorder prediction and structural coverage
 - Structural comparison of viral proteins
 - Bayesian network to predict viral-human PPIs
 - Functional enrichment analysis
 - Common and unique pathways within viral groups
 - Clustering viruses based on GSEA
 - Analysis of ZIKA interactome
 - Classification of Human Papillomaviruses (HPVs)
 - Evolutionary analysis on human genes

- Co-immunoprecipitation
- Webserver development
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.cell.2019.08.005>.

ACKNOWLEDGMENTS

We thank Jose Ignacio Garzon and Donald Petrey for their technical support. This work was funded by NIH grants 5R01GM109018-05 and 5U54CA209997-03 to S.D.S., 5R01GM030518-38 to B.H., 5R01GM117591 to R.R., and equipment grants S10OD012351 and S10OD021764 to the Department of Systems Biology.

AUTHOR CONTRIBUTIONS

Conceptualization, S.D.S.; Methodology, G.L., B.H., and S.D.S.; Validation, G.L., S.V.M., and E.R.W.; Formal Analysis and Investigation, G.L., K.P., J.A.P.-G., S.V.M., and E.R.W.; Funding Acquisition and Resources, B.H. and S.D.S.; Data Curation, G.L., T.C., and O.E.; Writing, G.L., B.H., and S.D.S.; Visualization, G.L. and S.D.S.; Supervision, R.R., B.H., and S.D.S.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: February 7, 2019

Revised: May 17, 2019

Accepted: August 2, 2019

Published: August 29, 2019

REFERENCES

- Abrescia, N.G., Bamford, D.H., Grimes, J.M., and Stuart, D.I. (2012). Structure unifies the viral universe. *Annu. Rev. Biochem.* *81*, 795–822.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* *215*, 403–410.
- Apweiler, R., Bairoch, A., Wu, C.H., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., et al. (2004). UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res.* *32*, D115–D119.
- Artaza-Irigaray, C., Flores-Miramontes, M.G., Olszewski, D., Magaña-Torres, M.T., López-Cardona, M.G., Leal-Herrera, Y.A., Piña-Sánchez, P., Jave-Suárez, L.F., and Aguilar-Lemarroy, A. (2017). Genetic variability in E6, E7 and L1 genes of Human Papillomavirus 62 and its prevalence in Mexico. *Infect. Agent. Cancer* *12*, 15.
- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al.; The Gene Ontology Consortium (2000). Gene ontology: tool for the unification of biology. *Nat. Genet.* *25*, 25–29.
- Bader, G.D., and Hogue, C.W. (2003). An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* *4*, 2.
- Bairoch, A. (2000). The ENZYME database in 2000. *Nucleic Acids Res.* *28*, 304–305.
- Baltimore, D. (1971). Expression of animal virus genomes. *Bacteriol. Rev.* *35*, 235–241.
- Barreiro, L.B., and Quintana-Murci, L. (2010). From evolutionary genetics to human immunology: how selection shapes host defence genes. *Nat. Rev. Genet.* *11*, 17–30.
- Batra, J., Hultquist, J.F., Liu, D., Shtanko, O., Von Dollen, J., Satkamp, L., Jang, G.M., Luthra, P., Schwarz, T.M., Small, G.I., et al. (2018). Protein Interaction Mapping Identifies RBBP6 as a Negative Regulator of Ebola Virus Replication. *Cell* *175*, 1917–1930.e13.
- Belinky, F., Nativ, N., Stelzer, G., Zimmerman, S., Iny Stein, T., Safran, M., and Lancet, D. (2015). PathCards: multi-source consolidation of human biological pathways. *Database (Oxford)* *2015*, bav006.
- Ben-Ari Fuchs, S., Lieder, I., Stelzer, G., Mazor, Y., Buzhor, E., Kaplan, S., Bogoch, Y., Plaschkes, I., Shitrit, A., Rappaport, N., et al. (2016). GeneAnalytics: An Integrative Gene Set Analysis Tool for Next Generation Sequencing, RNAseq and Microarray Data. *OMICS* *20*, 139–151.
- Berg, M., and Stenlund, A. (1997). Functional interactions between papillomavirus E1 and E2 proteins. *J. Virol.* *71*, 3853–3863.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank. *Nucleic Acids Res.* *28*, 235–242.
- Branca, M., Ciotti, M., Giorgi, C., Santini, D., Di Bonito, L., Costa, S., Benedetto, A., Bonifacio, D., Di Bonito, P., Paba, P., et al.; HPV-PathogenISS Study Group (2007). Up-regulation of proliferating cell nuclear antigen (PCNA) is closely associated with high-risk human papillomavirus (HPV) and progression of cervical intraepithelial neoplasia (CIN), but does not predict disease outcome in cervical cancer. *Eur. J. Obstet. Gynecol. Reprod. Biol.* *130*, 223–231.
- Budhu, A., Chen, Y., Kim, J.W., Forgues, M., Valerie, K., Harris, C.C., and Wang, X.W. (2007). Induction of a unique gene expression profile in primary human hepatocytes by hepatitis C virus core, NS3 and NS5A proteins. *Carcinogenesis* *28*, 1552–1560.
- Budowski-Tal, I., Nov, Y., and Kolodny, R. (2010). FragBag, an accurate representation of protein structure, retrieves structural neighbors from the entire PDB quickly and accurately. *Proc. Natl. Acad. Sci. USA* *107*, 3481–3486.
- Buitrago-Pérez, A., Garaulet, G., Vázquez-Carballo, A., Paramio, J.M., and García-Escudero, R. (2009). Molecular Signature of HPV-Induced Carcinogenesis: pRb, p53 and Gene Expression Profiling. *Curr. Genomics* *10*, 26–34.
- Bustamante, C.D., Fledel-Alon, A., Williamson, S., Nielsen, R., Hubisz, M.T., Glanowski, S., Tanenbaum, D.M., White, T.J., Sninsky, J.J., Hernandez, R.D., et al. (2005). Natural selection on protein-coding genes in the human genome. *Nature* *437*, 1153–1157.
- Butt, A.Q., Ahmed, S., Maratha, A., and Miggin, S.M. (2014). Retraction. *J. Biol. Chem.* *289*, 25474.
- Calderwood, M.A., Venkatesan, K., Xing, L., Chase, M.R., Vazquez, A., Holthaus, A.M., Ewence, A.E., Li, N., Hirozane-Kishikawa, T., Hill, D.E., et al. (2007). Epstein-Barr virus and virus human protein interaction maps. *Proc. Natl. Acad. Sci. USA* *104*, 7606–7611.
- Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* *25*, 1972–1973.
- Chan, Y.K., and Gack, M.U. (2016). A phosphomimetic-based mechanism of dengue virus to antagonize innate immunity. *Nat. Immunol.* *17*, 523–530.
- Chavali, P.L., Stojic, L., Meredith, L.W., Joseph, N., Nahorski, M.S., Sanford, T.J., Sweeney, T.R., Krishna, B.A., Hosmillo, M., Firth, A.E., et al. (2017). Neurodevelopmental protein Musashi-1 interacts with the Zika genome and promotes viral replication. *Science* *357*, 83–88.
- Chen, Y.F., and Xia, Y. (2019). Convergent perturbation of the human domain-resolved interactome by viruses and mutations inducing similar disease phenotypes. *PLoS Comput. Biol.* *15*, e1006762.
- Chen, G., He, C., Li, L., Lin, A., Zheng, X., He, E., and Skog, S. (2013). Nuclear TK1 expression is an independent prognostic factor for survival in pre-malignant and malignant lesions of the cervix. *BMC Cancer* *13*, 249.
- Chen, S., Zhang, J., Duan, L., Zhang, Y., Li, C., Liu, D., Ouyang, C., Lu, F., and Liu, X. (2014). Identification of HnRNP M as a novel biomarker for colorectal carcinoma by quantitative proteomics. *Am. J. Physiol. Gastrointest. Liver Physiol.* *306*, G394–G403.
- Chen, T.S., Petrey, D., Garzon, J.I., and Honig, B. (2015). Predicting peptide-mediated interactions on a genome-wide scale. *PLoS Comput. Biol.* *11*, e1004248.

- Cho, N.H., Kang, S., Hong, S., An, H.J., Choi, Y.H., Jeong, G.B., and Choi, H.K. (2006). Elevation of cyclin B1, active cdc2, and HuR in cervical neoplasia with human papillomavirus type 18 infection. *Cancer Lett.* *232*, 170–178.
- Chojnacki, M., and Melendy, T. (2018). The human papillomavirus DNA helicase E1 binds, stimulates, and confers processivity to cellular DNA polymerase epsilon. *Nucleic Acids Res.* *46*, 229–241.
- Collier, B., Goobar-Larsson, L., Sokolowski, M., and Schwartz, S. (1998). Translational inhibition in vitro of human papillomavirus type 16 L2 mRNA mediated through interaction with heterogenous ribonucleoprotein K and poly(rC)-binding proteins 1 and 2. *J. Biol. Chem.* *273*, 22648–22656.
- Cubie, H.A. (2013). Diseases associated with human papillomavirus infection. *Virology* *445*, 21–34.
- Daugherty, M.D., and Malik, H.S. (2012). Rules of engagement: molecular insights from host-virus arms races. *Annu. Rev. Genet.* *46*, 677–700.
- Davis, Z.H., Verschueren, E., Jang, G.M., Kleffman, K., Johnson, J.R., Park, J., Von Dollen, J., Maher, M.C., Johnson, T., Newton, W., et al. (2015). Global mapping of herpesvirus-host protein complexes reveals a transcription strategy for late genes. *Mol. Cell* *57*, 349–360.
- de Chassey, B., Meyniel-Schicklin, L., Aublin-Gex, A., Navratil, V., Chantier, T., André, P., and Lotteau, V. (2013). Structure homology and interaction redundancy for discovering virus-host protein interactions. *EMBO Rep.* *14*, 938–944.
- de Chassey, B., Navratil, V., Tafforeau, L., Hiet, M.S., Aublin-Gex, A., Agaogue, S., Meiffren, G., Pradezynski, F., Faria, B.F., Chantier, T., et al. (2008). Hepatitis C virus infection protein network. *Mol. Syst. Biol.* *4*, 230.
- Dietrich-Goetz, W., Kennedy, I.M., Levins, B., Stanley, M.A., and Clements, J.B. (1997). A cellular 65-kDa protein recognizes the negative regulatory element of human papillomavirus late mRNA. *Proc. Natl. Acad. Sci. USA* *94*, 163–168.
- Dinkel, H., Van Roey, K., Michael, S., Davey, N.E., Weatheritt, R.J., Born, D., Speck, T., Krüger, D., Grebnev, G., Kuban, M., et al. (2014). The eukaryotic linear motif resource ELM: 10 years and counting. *Nucleic Acids Res.* *42*, D259–D266.
- Doorbar, J., Quint, W., Banks, L., Bravo, I.G., Stoler, M., Broker, T.R., and Stanley, M.A. (2012). The biology and life-cycle of human papillomaviruses. *Vaccine* *30* (Suppl 5), F55–F70.
- Dosztányi, Z., Csizmók, V., Tompa, P., and Simon, I. (2005). The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *J. Mol. Biol.* *347*, 827–839.
- Eissenberg, J.C., Ayyagari, R., Gomes, X.V., and Burgers, P.M. (1997). Mutations in yeast proliferating cell nuclear antigen define distinct sites for interaction with DNA polymerase delta and DNA polymerase epsilon. *Mol. Cell. Biol.* *17*, 6367–6378.
- Emmott, E., Sorgeloos, F., Caddy, S.L., Vashist, S., Sosnovtsev, S., Lloyd, R., Heesom, K., Locker, N., and Goodfellow, I. (2017). Norovirus-Mediated Modification of the Translational Landscape via Virus and Host-Induced Cleavage of Translation Initiation Factors. *Mol. Cell. Proteomics* *16* (Suppl 1), S215–S229.
- Enard, D., Cai, L., Gwennap, C., and Petrov, D.A. (2016). Viruses are a dominant driver of protein adaptation in mammals. *eLife* *5*, e12469.
- Federhen, S. (2012). The NCBI Taxonomy database. *Nucleic Acids Res.* *40*, D136–D143.
- Finn, R.D., Coghill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., et al. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* *44* (D1), D279–D285.
- Frank, E., Hall, M.A., and Witten, I.H. (2016). The WEKA Workbench. Online Appendix for “Data Mining: Practical Machine Learning Tools and Techniques”, Fourth Edition (Morgan Kaufmann).
- Funk, J.O., Waga, S., Harry, J.B., Espling, E., Stillman, B., and Galloway, D.A. (1997). Inhibition of CDK activity and PCNA-dependent DNA replication by p21 is blocked by interaction with the HPV-16 E7 oncoprotein. *Genes Dev.* *11*, 2090–2100.
- Fuss, J., and Linn, S. (2002). Human DNA polymerase epsilon colocalizes with proliferating cell nuclear antigen and DNA replication late, but not early, in S phase. *J. Biol. Chem.* *277*, 8658–8666.
- Garamszegi, S., Franzosa, E.A., and Xia, Y. (2013). Signatures of pleiotropy, economy and convergent evolution in a domain-resolved map of human-virus protein-protein interaction networks. *PLoS Pathog.* *9*, e1003778.
- Garzón, J.I., Deng, L., Murray, D., Shapira, S., Petrey, D., and Honig, B. (2016). A computational interactome and functional annotation for the human proteome. *eLife* *5*, e18715.
- Geleziunas, R., Xu, W., Takeda, K., Ichijo, H., and Greene, W.C. (2001). HIV-1 Nef inhibits ASK1-dependent death signalling providing a potential mechanism for protecting the infected host cell. *Nature* *410*, 834–838.
- Germain, M.A., Chatel-Chaix, L., Gagne, B., Bonneil, E., Thibault, P., Pradezynski, F., de Chassey, B., Meyniel-Schicklin, L., Lotteau, V., Baril, M., et al. (2014). Elucidating novel hepatitis C virus-host interactions using combined mass spectrometry and functional genomics approaches. *Mol. Cell Proteomics* *13*, 184–203.
- Geuens, T., Bouhy, D., and Timmerman, V. (2016). The hnRNP family: insights into their role in health and disease. *Hum. Genet.* *135*, 851–867.
- González Plaza, J.J., Hulak, N., Kausova, G., Zhumadilov, Z., and Akilzhanova, A. (2016). Role of metabolism during viral infections, and crosstalk with the innate immune system. *Intractable Rare Dis. Res.* *5*, 90–96.
- Grant, A., Ponia, S.S., Tripathi, S., Balasubramaniam, V., Miorin, L., Sourisseau, M., Schwarz, M.C., Sánchez-Seco, M.P., Evans, M.J., Best, S.M., and García-Sastre, A. (2016). Zika Virus Targets Human STAT2 to Inhibit Type I Interferon Signaling. *Cell Host Microbe* *19*, 882–890.
- Grinstein, E., Wernet, P., Snijders, P.J., Rösl, F., Weinert, I., Jia, W., Kraft, R., Schewe, C., Schwabe, M., Hauptmann, S., et al. (2002). Nucleolin as activator of human papillomavirus type 18 oncogene transcription in cervical cancer. *J. Exp. Med.* *196*, 1067–1078.
- Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* *52*, 696–704.
- Halehalli, R.R., and Nagarajaram, H.A. (2015). Molecular principles of human virus protein-protein interactions. *Bioinformatics* *31*, 1025–1033.
- Hall, M.A. (1999). Correlation-based Feature Selection for Machine Learning. PhD thesis (The University of Waikato).
- Hallast, P., Maisano Delsler, P., Batini, C., Zadic, D., Rocchi, M., Schempp, W., Tyler-Smith, C., and Jobling, M.A. (2016). Great ape Y Chromosome and mitochondrial DNA phylogenies reflect subspecies structure and patterns of mating and dispersal. *Genome Res.* *26*, 427–439.
- Hamel, R., Dejamac, O., Wichit, S., Ekchariyawat, P., Neyret, A., Lupertlop, N., Perera-Lecoin, M., Surasombatpattana, P., Taligani, L., Thomas, F., et al. (2015). Biology of Zika Virus Infection in Human Skin Cells. *J. Virol.* *89*, 8880–8896.
- Han, X., Han, Y., Jiao, H., and Jie, Y. (2015). 14-3-3 ζ regulates immune response through Stat3 signaling in oral squamous cell carcinoma. *Mol. Cells* *38*, 112–121.
- Hu, S., Du, M.Q., Park, S.M., Alcivar, A., Qu, L., Gupta, S., Tang, J., Baens, M., Ye, H., Lee, T.H., et al. (2006). cIAP2 is a ubiquitin protein ligase for BCL10 and is dysregulated in mucosa-associated lymphoid tissue lymphomas. *J. Clin. Invest.* *116*, 174–181.
- Huang, W., Sherman, B.T., and Lempicki, R.A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* *37*, 1–13.
- Huang, W., Sherman, B.T., and Lempicki, R.A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* *4*, 44–57.
- Hughes, A.L., Packer, B., Welch, R., Bergen, A.W., Chanock, S.J., and Yeager, M. (2003). Widespread purifying selection at polymorphic sites in human protein-coding loci. *Proc. Natl. Acad. Sci. USA* *100*, 15754–15757.
- Hulo, C., de Castro, E., Masson, P., Bougueleret, L., Bairoch, A., Xenarios, I., and Le Mercier, P. (2011). ViralZone: a knowledge resource to understand virus diversity. *Nucleic Acids Res.* *39*, D576–D582.
- Hulsen, T., de Vlieg, J., and Alkema, W. (2008). BioVenn - a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genomics* *9*, 488.

- IARC. (2019). Agents Classified by the IARC Monographs, Volumes 1–124, <https://monographs.iarc.fr/agents-classified-by-the-iarc/>.
- Jäger, S., Cimermancic, P., Gulbahce, N., Johnson, J.R., McGovern, K.E., Clarke, S.C., Shales, M., Mercenne, G., Pache, L., Li, K., et al. (2011). Global landscape of HIV-human protein complexes. *Nature* **481**, 365–370.
- Johannessen, C.M., Boehm, J.S., Kim, S.Y., Thomas, S.R., Wardwell, L., Johnson, L.A., Emery, C.M., Stransky, N., Cogdill, A.P., Barretina, J., et al. (2010). COT drives resistance to RAF inhibition through MAP kinase pathway reactivation. *Nature* **468**, 968–972.
- Johansen, L.M., Brannan, J.M., Delos, S.E., Shoemaker, C.J., Stossel, A., Lear, C., Hoffstrom, B.G., Dewald, L.E., Schornberg, K.L., Scully, C., et al. (2013). FDA-approved selective estrogen receptor modulators inhibit Ebola virus infection. *Sci. Transl. Med.* **5**, 190ra79.
- Kajitani, N., and Schwartz, S. (2015). RNA Binding Proteins that Control Human Papillomavirus Gene Expression. *Biomolecules* **5**, 758–774.
- Karst, S.M., Wobus, C.E., Goodfellow, I.G., Green, K.Y., and Virgin, H.W. (2014). Advances in norovirus biology. *Cell Host Microbe* **15**, 668–680.
- Kasprzyk, A. (2011). BioMart: driving a paradigm change in biological data management. *Database (Oxford)* **2011**, bar049.
- Katzenellenbogen, R.A., Vliet-Gregg, P., Xu, M., and Galloway, D.A. (2010). Cytoplasmic poly(A) binding proteins regulate telomerase activity and cell growth in human papillomavirus type 16 E6-expressing keratinocytes. *J. Virol.* **84**, 12934–12944.
- Kelman, Z. (1997). PCNA: structure, functions and interactions. *Oncogene* **14**, 629–640.
- Kerrien, S., Aranda, B., Breuza, L., Bridge, A., Broackes-Carter, F., Chen, C., Duesbury, M., Dumousseau, M., Feuermann, M., Hinz, U., et al. (2012). The IntAct molecular interaction database in 2012. *Nucleic Acids Res.* **40**, D841–D846.
- Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797.
- Kuny, C.V., Chinchilla, K., Culbertson, M.R., and Kalejta, R.F. (2010). Cyclin-dependent kinase-like function is shared by the beta- and gamma- subset of the conserved herpesvirus protein kinases. *PLoS Pathog.* **6**, e1001092.
- Lee, D.K., Kim, B.C., Kim, I.Y., Cho, E.A., Satterwhite, D.J., and Kim, S.J. (2002). The human papilloma virus E7 oncoprotein inhibits transforming growth factor-beta signaling by blocking binding of the Smad complex to its target sequence. *J. Biol. Chem.* **277**, 38557–38564.
- Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44** (W1), W242–W245.
- Ley, K. (2003). The role of selectins in inflammation and disease. *Trends Mol. Med.* **9**, 263–268.
- Li, X., and Coffino, P. (1996). High-risk human papillomavirus E6 protein has two distinct binding sites within p53, of which only one determines degradation. *J. Virol.* **70**, 4509–4516.
- Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659.
- Li, X., Johansson, C., Glahder, J., Mossberg, A.K., and Schwartz, S. (2013). Suppression of HPV-16 late L1 5'-splice site SD3632 by binding of hnRNP D proteins and hnRNP A2/B1 to upstream AUAGUA RNA motifs. *Nucleic Acids Res.* **41**, 10488–10508.
- Li, C., Xu, D., Ye, Q., Hong, S., Jiang, Y., Liu, X., Zhang, N., Shi, L., Qin, C.F., and Xu, Z. (2016). Zika Virus Disrupts Neural Progenitor Development and Leads to Microcephaly in Mice. *Cell Stem Cell* **19**, 120–126.
- Lima, N.S., Rolland, M., Modjarrad, K., and Trautmann, L. (2017). T Cell Immunity and Zika Virus Vaccine Development. *Trends Immunol.* **38**, 594–605.
- Liu, H.M., Loo, Y.M., Horner, S.M., Zornetzer, G.A., Katze, M.G., and Gale, M., Jr. (2012). The mitochondrial targeting chaperone 14-3-3 σ regulates a RIG-I translocon that mediates membrane association and innate antiviral immunity. *Cell Host Microbe* **11**, 528–537.
- Löytynoja, A. (2014). Phylogeny-aware alignment with PRANK. *Methods Mol. Biol.* **1079**, 155–170.
- Lozier, M., Adams, L., Febo, M.F., Torres-Aponte, J., Bello-Pagan, M., Ryyf, K.R., Munoz-Jordan, J., Garcia, M., Rivera, A., Read, J.S., et al. (2016). Incidence of Zika Virus Disease by Age and Sex - Puerto Rico, November 1, 2015–October 20, 2016. *MMWR Morb. Mortal. Wkly. Rep.* **65**, 1219–1223.
- Luck, K., Sheynkman, G.M., Zhang, I., and Vidal, M. (2017). Proteome-Scale Human Interactomics. *Trends Biochem. Sci.* **42**, 342–354.
- Ma, Y., Walsh, M.J., Bernhardt, K., Ashbaugh, C.W., Trudeau, S.J., Ashbaugh, I.Y., Jiang, S., Jiang, C., Zhao, B., Root, D.E., et al. (2017). CRISPR/Cas9 Screens Reveal Epstein-Barr Virus-Transformed B Cell Host Dependency Factors. *Cell Host Microbe* **21**, 580–591.
- Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C.J., Lu, S., Chitsaz, F., Derbyshire, M.K., Geer, R.C., Gonzales, N.R., et al. (2017). CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* **45** (D1), D200–D203.
- Melendy, T., Sedman, J., and Stenlund, A. (1995). Cellular factors required for papillomavirus DNA replication. *J. Virol.* **69**, 7857–7867.
- Mihara, T., Nishimura, Y., Shimizu, Y., Nishiyama, H., Yoshikawa, G., Uehara, H., Hingamp, P., Goto, S., and Ogata, H. (2016). Linking Virus Genomes with Host Taxonomy. *Viruses* **8**, 66.
- Mirabello, L., Yeager, M., Yu, K., Clifford, G.M., Xiao, Y., Zhu, B., Cullen, M., Boland, J.F., Wentzensen, N., Nelson, C.W., et al. (2017). HPV16 E7 Genetic Conservation Is Critical to Carcinogenesis. *Cell* **170**, 1164–1174.
- Mohr, I.J., Clark, R., Sun, S., Androphy, E.J., MacPherson, P., and Botchan, M.R. (1990). Targeting the E1 replication protein to the papillomavirus origin of replication by complex formation with the E2 transactivator. *Science* **250**, 1694–1699.
- Müller, N., Avota, E., Schneider-Schaulies, J., Harms, H., Krohne, G., and Schneider-Schaulies, S. (2006). Measles virus contact with T cells impedes cytoskeletal remodeling associated with spreading, polarization, and CD3 clustering. *Traffic* **7**, 849–858.
- Muñoz, N., Bosch, F.X., de Sanjosé, S., Herrero, R., Castellsagué, X., Shah, K.V., Snijders, P.J., and Meijer, C.J.; International Agency for Research on Cancer Multicenter Cervical Cancer Study Group (2003). Epidemiologic classification of human papillomavirus types associated with cervical cancer. *N. Engl. J. Med.* **348**, 518–527.
- Musso, D., and Gubler, D.J. (2016). Zika Virus. *Clin. Microbiol. Rev.* **29**, 487–524.
- NCBI Resource Coordinators (2017). Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **45** (D1), D12–D17.
- Nicod, C., Banaei-Esfahani, A., and Collins, B.C. (2017). Elucidation of host-pathogen protein-protein interactions to uncover mechanisms of host cell rewiring. *Curr. Opin. Microbiol.* **39**, 7–15.
- Ohta, T., Michel, J.J., Schottelius, A.J., and Xiong, Y. (1999). ROC1, a homolog of APC11, represents a family of cullin partners with an associated ubiquitin ligase activity. *Mol. Cell* **3**, 535–541.
- Ouyang, P., Rakus, K., van Beurden, S.J., Westphal, A.H., Davison, A.J., Gatherer, D., and Vanderplasschen, A.F. (2014). IL-10 encoded by viruses: a remarkable example of independent acquisition of a cellular gene by viruses and its subsequent evolution in the viral genome. *J. Gen. Virol.* **95**, 245–262.
- Pang, C.L., Toh, S.Y., He, P., Teissier, S., Ben Khalifa, Y., Xue, Y., and Thierry, F. (2014). A functional interaction of E7 with B-Myb-MuvB complex promotes acute cooperative transcriptional activation of both S- and M-phase genes. (129 c). *Oncogene* **33**, 4039–4049.
- Petrey, D., Xiang, Z., Tang, C.L., Xie, L., Gimpelev, M., Mitros, T., Soto, C.S., Goldsmith-Fischman, S., Kernysky, A., Schlessinger, A., et al. (2003). Using multiple structure alignments, fast model building, and energetic analysis in fold recognition and homology modeling. *Proteins* **53** (Suppl 6), 430–435.
- Pichlmair, A., Kandasamy, K., Alvisi, G., Mulhern, O., Sacco, R., Habjan, M., Binder, M., Stefanovic, A., Eberle, C.A., Goncalves, A., et al. (2012). Viral immune modulators perturb the human molecular network by common and unique strategies. *Nature* **487**, 486–490.
- Pickett, B.E., Sadat, E.L., Zhang, Y., Noronha, J.M., Squires, R.B., Hunt, V., Liu, M., Kumar, S., Zaremba, S., Gu, Z., et al. (2012). ViPR: an open bioinformatics database and analysis resource for virology research. *Nucleic Acids Res.* **40**, D593–D598.

- Pietsch, E.C., and Murphy, M.E. (2008). Low risk HPV-E6 traps p53 in the cytoplasm and induces p53-dependent apoptosis. *Cancer Biol. Ther.* *7*, 1916–1918.
- Pyeon, D., Newton, M.A., Lambert, P.F., den Boon, J.A., Sengupta, S., Marsit, C.J., Woodworth, C.D., Connor, J.P., Haugen, T.H., Smith, E.M., et al. (2007). Fundamental differences in cell cycle deregulation in human papillomavirus-positive and human papillomavirus-negative head/neck and cervical cancers. *Cancer Res.* *67*, 4605–4619.
- Quinn, K., Brindley, M.A., Weller, M.L., Kaludov, N., Kondratowicz, A., Hunt, C.L., Sinn, P.L., McCray, P.B., Jr., Stein, C.S., Davidson, B.L., et al. (2009). Rho GTPases modulate entry of Ebola virus and vesicular stomatitis virus pseudotyped vectors. *J. Virol.* *83*, 10176–10186.
- R Development Core Team (2016). R: A language and environment for statistical computing (R Foundation for Statistical Computing).
- Ramage, H.R., Kumar, G.R., Verschuere, E., Johnson, J.R., Von Dollen, J., Johnson, T., Newton, B., Shah, P., Horner, J., Krogan, N.J., et al. (2015). A combined proteomics/genomics approach links hepatitis C virus infection with nonsense-mediated mRNA decay. *Mol. Cell* *57*, 329–340.
- Raman, R., Tharakaraman, K., Sasisekharan, V., and Sasisekharan, R. (2016). Glycan-protein interactions in viral pathogenesis. *Curr. Opin. Struct. Biol.* *40*, 153–162.
- Remmert, M., Biegert, A., Hauser, A., and Söding, J. (2011). HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat. Methods* *9*, 173–175.
- Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* *16*, 276–277.
- Rogers, J., and Gibbs, R.A. (2014). Comparative primate genomics: emerging patterns of genome content and dynamics. *Nat. Rev. Genet.* *15*, 347–359.
- Rose, A.S., Bradley, A.R., Valasatava, Y., Duarte, J.M., Pricl, A., and Rose, P.W. (2018). NGL viewer: web-based molecular graphics for large complexes. *Bioinformatics* *34*, 3755–3758.
- Sawyer, S.L., Wu, L.I., Emerman, M., and Malik, H.S. (2005). Positive selection of primate TRIM5alpha identifies a critical species-specific retroviral restriction domain. *Proc. Natl. Acad. Sci. USA* *102*, 2832–2837.
- Scaturro, P., Stukalov, A., Haas, D.A., Cortese, M., Draganova, K., Plaszczyca, A., Bartenschlager, R., Götz, M., and Pichlmair, A. (2018). An orthogonal proteomic survey uncovers novel Zika virus host factors. *Nature* *561*, 253–257.
- Shah, P.S., Link, N., Jang, G.M., Sharp, P.P., Zhu, T., Swaney, D.L., Johnson, J.R., Von Dollen, J., Ramage, H.R., Satkamp, L., et al. (2018). Comparative Flavivirus-Host Protein Interaction Mapping Reveals Mechanisms of Dengue and Zika Virus Pathogenesis. *Cell* *175*, 1931–1945.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* *13*, 2498–2504.
- Subbiah, S., Laurents, D.V., and Levitt, M. (1993). Structural similarity of DNA-binding domains of bacteriophage repressors and the globin core. *Curr. Biol.* *3*, 141–148.
- Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* *102*, 15545–15550.
- Sun, H., Liu, T., Zhu, D., Dong, X., Liu, F., Liang, X., Chen, C., Shao, B., Wang, M., and Wang, Y. (2017). HnRNPM and CD44s expression affects tumor aggressiveness and predicts poor prognosis in breast cancer with axillary lymph node metastases. *Genes Chromosomes Cancer* *56*, 598–607.
- Takizawa, C.G., and Morgan, D.O. (2000). Control of mitosis by changes in the subcellular location of cyclin-B1-Cdk1 and Cdc25C. *Curr. Opin. Cell Biol.* *12*, 658–665.
- Tang, H., Hammack, C., Ogden, S.C., Wen, Z., Qian, X., Li, Y., Yao, B., Shin, J., Zhang, F., Lee, E.M., et al. (2016). Zika Virus Infects Human Cortical Neural Progenitors and Attenuates Their Growth. *Cell Stem Cell* *18*, 587–590.
- Tanner, J., Whang, Y., Sample, J., Sears, A., and Kieff, E. (1988). Soluble gp350/220 and deletion mutant glycoproteins block Epstein-Barr virus adsorption to lymphocytes. *J. Virol.* *62*, 4452–4464.
- The Gene Ontology Consortium (2017). Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res.* *45* (D1), D331–D338.
- Tjalma, W.A., Weyler, J.J., Bogers, J.J., Pollefliet, C., Baay, M., Goovaerts, G.C., Vermorken, J.B., van Dam, P.A., van Marck, E.A., and Buytaert, P.M. (2001). The importance of biological factors (bcl-2, bax, p53, PCNA, MI, HPV and angiogenesis) in invasive cervical cancer. *Eur. J. Obstet. Gynecol. Reprod. Biol.* *97*, 223–230.
- Tripathi, S., Pohl, M.O., Zhou, Y., Rodriguez-Frandsen, A., Wang, G., Stein, D.A., Moulton, H.M., DeJesus, P., Che, J., Mulder, L.C., et al. (2015). Meta- and Orthogonal Integration of Influenza “OMICS” Data Defines a Role for UBR4 in Virus Budding. *Cell Host Microbe* *18*, 723–735.
- Tsuchiya, N., Kamei, D., Takano, A., Matsui, T., and Yamada, M. (1998). Cloning and characterization of a cDNA encoding a novel heterogeneous nuclear ribonucleoprotein-like protein and its expression in myeloid leukemia cells. *J. Biochem.* *123*, 499–507.
- Uetz, P., Dong, Y.A., Zeretzke, C., Atzler, C., Baiker, A., Berger, B., Rajagopala, S.V., Roupelieva, M., Rose, D., Fossum, E., et al. (2006). Herpesviral protein networks and their interaction with the human proteome. *Science* *311*, 239–242.
- Van den Broeke, C., Jacob, T., and Favoreel, H.W. (2014). Rho’ing in and out of cells: viral interactions with Rho GTPase signaling. *Small GTPases* *5*, e28318.
- Vanhee, P., Reumers, J., Stricher, F., Baeten, L., Serrano, L., Schymkowitz, J., and Rousseau, F. (2010). PepX: a structural database of non-redundant protein-peptide complexes. *Nucleic Acids Res.* *38*, D545–D551.
- von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S.G., Fields, S., and Bork, P. (2002). Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* *417*, 399–403.
- Warnes, G.R., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., et al. (2016). gplots: Various R Programming Tools for Plotting Data.
- White, E.A., and Munger, K. (2017). Crowd Control: E7 Conservation Is the Key to Cancer. *Cell* *170*, 1057–1059.
- Wobus, C.E., Thackray, L.B., and Virgin, H.W., 4th. (2006). Murine norovirus: a model system to study norovirus biology and pathogenesis. *J. Virol.* *80*, 5104–5112.
- Wu, K.Y., Zuo, G.L., Li, X.F., Ye, Q., Deng, Y.Q., Huang, X.Y., Cao, W.C., Qin, C.F., and Luo, Z.G. (2016). Vertical transmission of Zika virus targeting the radial glial cells affects cortex development of offspring mice. *Cell Res.* *26*, 645–654.
- Xiang, Z., and Honig, B. (2001). Extending the accuracy limits of prediction for side-chain conformations. *J. Mol. Biol.* *311*, 421–430.
- Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* *24*, 1586–1591.
- Yang, A.S., and Honig, B. (2000a). An integrated approach to the analysis and modeling of protein sequences and structures. I. Protein structural alignment and a quantitative measure for protein structural distance. *J. Mol. Biol.* *307*, 665–678.
- Yang, A.S., and Honig, B. (2000b). An integrated approach to the analysis and modeling of protein sequences and structures. III. A comparative study of sequence conservation in protein structural families using multiple structural alignments. *J. Mol. Biol.* *307*, 691–711.
- Zhang, Q.C., Petrey, D., Deng, L., Qiang, L., Shi, Y., Thu, C.A., Bisikirska, B., Lefebvre, C., Accilli, D., Hunter, T., et al. (2012). Structure-based prediction of protein-protein interactions on a genome-wide scale. *Nature* *490*, 556–560.

STAR★METHODS

KEY RESOURCE TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse anti-flavivirus D1-4G2-4-15	EMD millipore	Cat#MAB10216; RRID: AB_827205
Goat anti-mouse IgG HRP-conjugated antibody	R&D Systems	Cat#HAF007; RRID: AB_357234
rabbit polyclonal anti-HA antibody	Sigma-Aldrich	Cat#H6908; RRID: AB_260070
rabbit anti-FLAG antibody	Cell Signaling Technology	Cat#2368; RRID: AB_2217020
anti-V5-HRP antibody	Life Technologies	Cat#R961-25; RRID:AB_2556565
c-Myc polyclonal antibody	Santa Cruz Biotechnology	Cat#SC-40; RRID:AB_627268
Goat anti-rabbit HRP-conjugated secondary antibody	Thermo Scientific	Cat#PI31460
Bacterial and Virus Strains		
Zika virus MR-766	BEI Resources	Cat#NR-50065
Chemicals, Peptides, and Recombinant Proteins		
Human IFN β 1a	PBL Assay Science	Cat#11410-2
Critical Commercial Assays		
VIP peroxidase substrate	Vector Laboratories	Cat#SK-4600
TransIT-LT1 Reagent	Mirus	Cat#MIR2300
Lipofectamine RNAiMAX Transfection Reagent	Invitrogen, Life Technologies	Cat#13778-030
High Capacity cDNA Reverse Transcription Kit	Applied Biosystems	Cat#43-688-14
iTaq Universal SYBR Green Supermix	Bio-Rad Laboratories	Cat#1725121
protein G magnetic beads, SureBeads	Bio-Rad Laboratories	Cat#1614023
SuperSignal West Femto Maximum Sensitivity Substrate	Thermo Scientific	Cat#PI34096
Deposited Data		
Protein modeling, prediction of PPIs, functional annotation	P-HIPSTER	http://phipster.org/
Evolutionary analyses	Github	https://github.com/RabadanLab/palmer
Experimental Models: Cell Lines		
293T cells	ATCC	Cat#CRL-3216
Vero cells	ATCC	Cat#CCL-81
Human breast adenocarcinoma cells (MCF-7)	ATCC	Cat#HTB-22
Oligonucleotides		
siRNA targeting human ESR1 SMARTpool	Dharmacon	Cat#L-003401-00-0005
AllStars Negative Control siRNA	QIAGEN	Cat#SI03650318
AllStars Hs Cell Death siRNA	QIAGEN	Cat#SI04381048
Recombinant DNA		
Plasmid: Human AMBP	The CCSB Human ORFeome collection 8.1	#Cat13590
Plasmid: Human BIRC2	The CCSB Human ORFeome collection 8.1	#Cat200
Plasmid: Human BRCA1 (2100 nt)	The CCSB Human ORFeome collection 8.1	#Cat14613
Plasmid: Human BRCA1 (4065 nt)	The CCSB Human ORFeome collection 8.1	#Cat55703
Plasmid: Human DOCK5	The CCSB Human ORFeome collection 8.1	#Cat13488
Plasmid: Human ESR1	The CCSB Human ORFeome collection 8.1	#Cat56919
Plasmid: Human ING2	The CCSB Human ORFeome collection 8.1	#Cat12445
Plasmid: Human ISG15	The CCSB Human ORFeome collection 8.1	#Cat6045
Plasmid: Human RABGGTA	The CCSB Human ORFeome collection 8.1	#Cat3713

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Plasmid: Human SFN	The CCSB Human ORFeome collection 8.1	#Cat2433
Plasmid: Human SPINT2	The CCSB Human ORFeome collection 8.1	#Cat550
Plasmid: Human STAC2	The CCSB Human ORFeome collection 8.1	#Cat55919
Plasmid: Human STAT3	The CCSB Human ORFeome collection 8.1	#Cat7727
Plasmid: Human STK32A	The CCSB Human ORFeome collection 8.1	#Cat7425
Plasmid: Human TP53	The CCSB Human ORFeome collection 8.1	#Cat3774
Plasmid: Human VARS	The CCSB Human ORFeome collection 8.1	#Cat6043
Plasmid: Human VAV1	The CCSB Human ORFeome collection 8.1	#Cat6612
Plasmid: Human VCAM1	The CCSB Human ORFeome collection 8.1	#Cat14543
Plasmid: Human SPINLW1 (EPPIN)	The CCSB Human ORFeome collection 8.1	#Cat11827
Plasmid: Human ABL1 pDONR223-ABL1	William Hahn & David Root; Johannessen et al., 2010	Addgene 23939
Plasmid: Human APP pLX304-APP	The CCSB Human ORFeome collection 8.1	Cat#8905
Plasmid: Human BIRC3 3xMyc-clAP2/pRK5	Xiaolu Yang; Hu et al., 2006	Addgene 27973
Plasmid: Human CUL1 pcDNA3-myc3-CUL1	Yue Xiong; Ohta et al., 1999	Addgene 19896
Plasmid: HCV NS5A pCMV-Tag1-NS5A	Xin Wang; Budhu et al., 2007	Addgene 17646
Plasmid: HSV-1 UL13 HSV-1 UL13	Robert Kalejta; Kuny et al., 2010	Addgene 26697
Plasmid: HIV-1 Nef (HIV1gp9) pCI NL4-3 Nef-HA-WT	Warner Greene	Addgene 24162
Plasmid: ZIKV MR-766 Capsid pLV_Zika_Cv_Flag	Vaithi Arumugaswami; unpublished	Addgene 79628
Plasmid: ZIKV MR-766 NS3 pLV_Zika_NS3_Flag	Vaithi Arumugaswami; unpublished	Addgene 79635
Plasmid: ZIKV MR-766 NS5 pLV_Zika_NS5_Flag	Vaithi Arumugaswami; unpublished	Addgene 79639
Ska	Yang and Honig, 2000a	http://honig.c2b2.columbia.edu/ska
CD-HIT	Li and Godzik, 2006	http://weizhongli-lab.org/cd-hit/
EMBOSS	Rice et al., 2000	http://emboss.sourceforge.net/
PrePPI	Zhang et al., 2012	http://honig.c2b2.columbia.edu/preppi
Software and Algorithms		
BLAST	Altschul et al., 1990	https://blast.ncbi.nlm.nih.gov/Blast.cgi
HHblits	Remmert et al., 2011	https://github.com/soedinglab/hh-suite
NEST	Petrey et al., 2003	http://honig.c2b2.columbia.edu/nest
IUPred	Dosztányi et al., 2005	http://iupred.elte.hu/
DAVID	Huang et al., 2009b	https://david.ncifcrf.gov/
GSEA	Subramanian et al., 2005	http://software.broadinstitute.org/gsea/index.jsp
PathCards	Belinky et al., 2015	https://pathcards.genecards.org/
BioVenn	Hulsen et al., 2008	http://www.biovenn.nl/
R	R Development Core Team, 2016	https://www.r-project.org/
Gplots package	Warnes et al., 2016	https://cran.r-project.org/web/packages/gplots/index.html
iTOL	Letunic and Bork, 2016	https://itol.embl.de/
Weka	Frank et al., 2016	https://www.cs.waikato.ac.nz/~ml/weka/
Biomart	Kasprzyk, 2011	http://useast.ensembl.org/biomart/martview/47aaecf5a2236f19252cc3c2ae91e094
Cytoscape	Shannon et al., 2003	https://cytoscape.org/
MCODE	Bader and Hogue, 2003	http://baderlab.org/Software/MCODE
geneAnalytics	Ben-Ari Fuchs et al., 2016	https://geneanalytics.genecards.org/

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Prism	GraphPad	https://www.graphpad.com/
Prank	Löytynoja, 2014	http://wasabiapp.org/software/prank/
TrimA1	Capella-Gutierrez et al., 2009	http://trimal.cgenomics.org/
PhyML	Guindon and Gascuel, 2003	http://www.atgc-montpellier.fr/phyml/
PAML	Yang, 2007	http://abacus.gene.ucl.ac.uk/software/paml.html
Other		
Database: virus-hostDB	Mihara et al., 2016	https://www.genome.jp/virushostdb
Database: NCBI	NCBI Resource Coordinators, 2017	https://www.ncbi.nlm.nih.gov/
Database: Uniprot	Altschul et al., 1990	https://www.uniprot.org/
Database: Conserved Domain Database	Marchler-Bauer et al., 2017	https://www.ncbi.nlm.nih.gov/cdd/
Database: PDB	Berman et al., 2000	http://www.rcsb.org/
Database: PepX	Vanhee et al., 2010	http://www.switchlab.org/bioinformatics/pepx
Database: ELM	Dinkel et al., 2014	http://elm.eu.org/
Database: Intact	Kerrien et al., 2012	https://www.ebi.ac.uk/intact/
Database: Pfam	Finn et al., 2016	https://pfam.xfam.org/
Database: PISA	Krissinel and Henrick, 2007	http://www.ebi.ac.uk/pdbe/pisa/

See [Table S8](#) for a description of viral and human plasmids utilized in this work as well as primers employed for Co-immunoprecipitation assays.

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for reagents may be directed to, and will be fulfilled by the corresponding author Sagi D. Shapira (ss4197@cumc.columbia.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS**Cell lines, virus strains**

293T, Vero and MCF7 cells were maintained in Dulbecco's modified Eagle's medium (DMEM; GIBCO, Life Technologies) supplemented with 10% fetal bovine serum (FBS; Hyclone Laboratories) and 1% penicillin and streptomycin (GIBCO, Life Technologies) at 37°C and 5% CO₂. All cell lines have been authenticated by ATCC. Additionally, each cell line was inspected through microscopy for morphology and growth characteristics. Cell lines were tested for mycoplasma contamination and found to be negative. MCF7 cells are female cells as described by ATCC. Zika virus (ZIKV), MR 766 strain obtained from BEI Resources was kindly provided by Dr. Vincent Racaniello (Columbia University Medical Center), was amplified once on Vero cells. Virus titer was determined by focus-forming assay on Vero cells.

METHOD DETAILS**Assembly of viral protein and human protein dataset**

We compiled a viral protein dataset containing 12,237 viral proteins corresponding to 1,001 completely sequenced human viruses from virus-hostDB (Mihara et al., 2016) as of October, 2016 (Table S1; <https://www.genome.jp/virushostdb>). Originally, the dataset contained 12,520 viral proteins and 1,028 human viruses. The dataset was manually curated by cross-referencing viral taxonomy IDs to other viral databases and viruses with poorly annotated genomes (viruses with < 4 proteins and a genomic coverage < 80%) and viruses wrongly annotated as human viruses were discarded (Federhen, 2012; Hulo et al., 2011; Pickett et al., 2012). Information for each virus (nucleic type, taxonomic classification, taxonomic identifier) and viral protein (protein database identifiers, description and amino acid sequence) was retrieved from virus-hostDB (Mihara et al., 2016). Additionally, proteins annotated as polypeptides in virus-hostDB were parsed into individual proteins, whenever possible, using the annotation of proteins and mature peptides as described in the NCBI protein database (NCBI Resource Coordinators, 2017). We assigned Uniprot accession codes (ACs) to viral proteins by Blasting each protein sequence against the Uniprot database (Altschul et al., 1990; Apweiler et al., 2004) considering only those hits

with maximum E-value of 10^{-20} , a minimum alignment coverage of 70% of the shortest sequence aligned and a minimum sequence identity of 90%. The human protein dataset contains 20,113 non-immunoglobulin human proteins obtained from the Uniprot database as described by Garzon et. al (Apweiler et al., 2004; Garzón et al., 2016).

Protein modeling, structural neighbor search

The P-HIPSTer (Pathogen-Host Interactome Prediction using SStructurE similaRity) algorithm enables systematic interrogation of pan-human virus interactions by exploiting both sequence- and structure-based information from atomic structures taken from the PDB, and from homology models, to account for both domain-domain and peptide-domain interactions. Three-dimensional models for full-length viral proteins and protein domains, as defined by the Conserved Domain Database (Marchler-Bauer et al., 2017), are either taken directly from the PDB (Berman et al., 2000) or built by homology modeling. The homology pipeline described in Garzón et al. (2016) was adapted to model large polyproteins and to account for potential conformational variability whenever possible by considering multiple templates for homology modeling rather than only considering the best scoring template. Template search is performed in three steps where each step is run only if the preceding step reports no templates. Templates are identified based on the significance of the sequence alignment between the query protein and the matching protein structure, requiring an E-value $< 10^{-12}$ and some additional criteria imposed at each step. The first step runs one iteration of Blast (Altschul et al., 1990) against the PDB database and identifies the set of non-overlapping templates (matching different segments of the query protein sequence) with the lowest E-value. Overlapping templates with higher E-values are also considered for modeling only if their E-values are $< 10^{-12}$ and lie within 10^{-10} of the E-value reported for the best hit found in that particular protein segment. The second and third step runs HHblits (Remmert et al., 2011) and five iterations of Blast (Altschul et al., 1990) respectively against the PDB database, identifying templates in a similar manner with the exception that overlapping templates with higher E-values are considered only if their E-values are $< 10^{-12}$ and lie within 10^{-2} of the E-value reported for the best hit found in that particular protein segment. Atomic models were built with NEST (Petrey et al., 2003; Xiang and Honig, 2001) based on the alignments provided either by Blast or by HHblits. Homology modeling of human proteins was carried out as described by Garzón et al. (2016). In modeling the set of viral proteins used in this study, we found that the greatest structural coverage was obtained for proteomes of (+)ssRNA and (-)ssRNA viruses may reflect the tendency of their proteomes to have fewer disordered regions (Figure S1C), or the overrepresentation of viral proteins from RNA viruses in the PDB (that serve as templates for structural modeling; 4,791, 3,104 and 2,638 viral proteins from RNA, DNA and RT viruses respectively as of April, 2017).

The structure of the protein complexes used as templates for interaction models are defined by i) PDB 'biounit', likely representing the biological relevant quaternary structure for interacting proteins; ii) the PISA database of predicted biounits (Krissinel and Henrick, 2007) and; iii) the PDB file (Garzón et al., 2016). In order to identify structurally similar proteins, a structural neighbor search is carried out for each human and viral protein or domain (either taken directly from the PDB or modeled by homology) against the PDB database with Ska (Petrey et al., 2003; Yang and Honig, 2000b) using a Protein Structural Distance (PSD) ≤ 0.6 .

Disorder prediction and structural coverage

The fraction of disordered residues and the structural coverage for each Baltimore category (Baltimore, 1971) is calculated at the residue level for each corresponding virus. Disorder prediction is calculated with IUPred (Dosztányi et al., 2005), where each residue is predicted as disordered if its IUPred score is > 0.5 . For each virus within a Baltimore category we calculate the fraction of residues predicted as disordered. Similarly, we define structural coverage for each virus as the fraction of residues within a virus present in an X-ray structure or homology model.

Structural comparison of viral proteins

All-against-all structural comparison of atomic structures (either modeled or derived from the PDB), corresponding to viral capsid proteins or to a subset of viral proteins with low pairwise sequence identity, was performed using Ska (Yang and Honig, 2000b). Structural similarity between viral capsids has been used to infer functional and evolutionary relationships where genomic similarities are no longer observable (Abrescia et al., 2012). Here, we extend the same principle to a subset of viral capsids whose structure are modeled by homology. Similarly, structural comparison of viral proteins with low pairwise sequence identity permit us to identify potential functional relationships between pairs of viral proteins that are not detectable using sequence alone. A low sequence similarity subset was assembled at 40% sequence identity with CD-HIT (Li and Godzik, 2006). Considering that most viral proteins are described by several protein structures (corresponding to either the full sequence or domains and additionally modeled using multiple templates), we define the structural distance between two proteins as the shortest distance reported between any two protein structures, where each structure corresponds to one of the two proteins. We use both the protein structural distance (PSD) and the structural alignment score (SAS) to measure distance between any two atomic structures (Subbiah et al., 1993; Yang and Honig, 2000a). PSD reflects local and/or global structural similarity and can capture structural relationships and functional relationships involving only a small number of secondary structure elements (> 3). SAS, on the other hand, takes into account the length of the alignment and better reflects global structural similarity. By combining both PSD and SAS we expect to capture both local and global structural relationships between pairs of proteins. Two atomic structures are considered structurally similar when their PSD is ≤ 0.6 and their SAS is ≤ 3.5 Å: while the PSD cutoff of 0.6 indicates that there is good structural similarity over at least part of both structures, an SAS cutoff of 3.5 Å better reflects a global similarity (Budowski-Tal et al., 2010; Zhang

et al., 2012). Pairwise sequence identities described in Figure S5 were computed with the Needleman-Wunsch algorithm implemented in the EMBOSS package (Rice et al., 2000).

Bayesian network to predict viral-human PPIs

P-HIPSTER adapts the Bayesian network underlying the PrePPI (Garzón et al., 2016; Zhang et al., 2012) algorithm and reports a likelihood ratio (LR) of the interaction between any pathogen and host protein using three separate structure-based evidences E_x : i) the structural modeling evidence (E_{dom}) evaluates the potential direct interaction between two query proteins through their folded domains (Zhang et al., 2012); ii) the protein-peptide evidence (E_{pep}) evaluates the likelihood of a PPI through an unstructured domain and a folded domain (Chen et al., 2015); iii) the partner redundancy evidence (E_{redu}) infers a likelihood of a PPI based on the number of structural neighbors of one query protein known to interact with the remaining query protein (de Chasseay et al., 2013; Garzón et al., 2016). E_{dom} evaluates the potential interaction between two folded protein domains based on four criteria (Zhang et al., 2012): i) SIM: the average structural similarity of the query domain pair to the interacting subunits in an interaction template (an experimentally solved protein complex involving two proteins); ii) SIZ: the number of interacting residue pairs in the interaction template that are structurally aligned to residue pairs in the query domains after structural superposition; iii) COV: the fraction of interacting residue pairs in the interaction template that are structurally aligned to residue pairs in the query domains after structural superposition and; iv) OS: same as SIZ, with the additional condition that each residue in the interacting pair aligns to a residue predicted to be interfacial in the query domains. E_{pep} (Chen et al., 2015) reports the maximum LR assessing the potential interaction between an unstructured domain and a folded domain in two independent evidences: i) PepX predicts peptide-domain interactions based on experimentally determined complexes (Vanhee et al., 2010) using both structural and sequence similarity; ii) PepELM predicts peptide-domain interactions based on the Eukaryotic Linear Motif (ELM) database (Dinkel et al., 2014), by identifying a PFAM sequence signature in one query protein and a motif sequence signature in the remaining query protein, both mapping to the same ELM class. E_{redu} does not necessarily imply a direct physical interaction between query proteins but it is used to further support the predictions reported by E_{dom} and/or E_{pep} . This evidence combines knowledge on PPIs extracted from PPI databases with structural similarity: a Likelihood Ratio (LR) of the interaction between two proteins A and B is assigned based on the number of structural neighbors of A known to interact with B, or viceversa, and the number of structural neighbors of A and B that are known to interact (Garzón et al., 2016).

Training was carried out on the yeast interactome (Garzón et al., 2016). The scores for each evidence E are partitioned into n bins b_1, b_2, \dots, b_n . A likelihood ratio $LR_{b_i}^{E_x}$ is assigned to each bin. $LR_{b_i}^{E_x}$ corresponds to the percentage of protein pairs in a positive gold standard dataset of PPIs in yeast with a score for E_x in bin b_i divided by the percentage of protein pairs in the negative gold standard dataset of PPIs in yeast with a score for E_x in the same bin b_i . Details regarding the Bayesian network training and the calculation of LR scores associated to each evidence have been described (Chen et al., 2015; Garzón et al., 2016; Zhang et al., 2012). P-HIPSTER integrates the LR scores obtained from each structural evidence E_x with a Naive Bayes approach. Since E_{dom} and E_{pep} evidences can be considered as mutually exclusive, we only consider the maximum LR reported by either of them:

$$LR_{P-HIPSTER} = \max(LR^{E_{dom}}, LR^{E_{pep}}) \times LR^{E_{redu}} \quad (1)$$

Overall statistics of the set of predictions (number of predicted PPIs per viral protein, per virus and per host proteins) are computed using only instances for which $LR_{P-HIPSTER}$ is equal or higher than an LR threshold of 100 (7,463 viral proteins, 990 viruses and 5,749 human proteins).

Functional enrichment analysis

We used DAVID to identify the set of biological pathways and molecular functions enriched within the subset of the human proteome predicted to interact with human viruses with an $LR_{P-HIPSTER} \geq 100$ (Huang et al., 2009a, 2009b). We used two different subsets: i) a set of 5,749 human proteins predicted to interact with at least one human virus with an $LR_{P-HIPSTER} \geq 100$ and; ii) a set of 173 human proteins predicted to interact with ≥ 100 human viruses with an $LR_{P-HIPSTER} \geq 100$. We corrected the background of the enrichment analysis for the 20,113 non-immunoglobulin human proteins considered by P-HIPSTER in this study.

Similar to previous work, we use the ranked list of predicted human interactors for a given viral protein to identify the biological pathways enriched (Garzón et al., 2016). Enrichment of gene sets for viral proteins are computed with GSEA software from mSigDB (Subramanian et al., 2005). For each viral protein, a ranked list of predicted interacting human proteins is constructed based on the $LR_{P-HIPSTER}$. In order to account for background and potential technical bias, we subtracted the corresponding average LR derived from the predicted PrePPI human interactome to the $LR_{P-HIPSTER}$ (Garzón et al., 2016). Enrichment was calculated considering only the rank order (classic mode). Gene sets were derived from PathCards (Belinky et al., 2015) and considered enriched if the reported q -value was < 0.01 . In order to collect the set of enriched pathways for any given virus, we computed the union of enriched gene sets (q -value < 0.01) of the corresponding viral proteins.

Common and unique pathways within viral groups

The discovery of common and unique pathways predicted to be targeted by DNA, RNA and RT viruses is carried out using a two-step procedure where first the most relevant pathways are selected for each viral category (DNA, RNA or RT) and then the overlap between

the different categories (defined by their set of relevant pathways) is computed. Out of the 1,001 human viruses considered in our dataset, we excluded 148 viruses with low protein structural coverage (> 50% of their residues were not successfully modeled). In order to reduce potential bias by larger viral families within a category we consider a pathway relevant for a particular category whenever it is found enriched ($q\text{-val} < 0.01$) in at least of 50% of the viruses within a viral family in at least 50% of the viral families. Venn diagrams describing the overlap between different nucleic acid type viral groups were computed with BioVenn and manually edited for visualization purposes (Hulsen et al., 2008). The set pathways within each region of the Venn diagram were manually classified into 26 different categories: apoptosis, cell adhesion, cell cycle, cell differentiation, cell growth, cell organization, cellular trafficking, development, disease, DNA, drug, extracellular matrix, immune, lipid, metabolism, microRNA, mRNA processing, pathogen, protein processing, RNA, signaling, transcription, translation, transport, viral and/or other. The significance of each category found within the different regions of the Venn diagram was assessed with a permutation test.

Clustering viruses based on GSEA

In order to avoid potential clustering errors due to technical noise (e.g., unrelated viruses with only few modeled proteins can cluster together leading to wrong conclusions) we applied a stringent filter where we only consider 568 human viruses with: i) annotated mature viral proteins where $\geq 50\%$ of their proteins and their residues can be modeled and; ii) a number of viral proteins that is consistent with other viruses within the same genus or family (removing viruses with polyproteins not annotated as such; e.g., noroviruses with three viral proteins while the majority of noroviruses have eight viral proteins). For each virus, we calculated the union of enriched biological pathways with a $q\text{-value} < 0.01$ using the set of enriched pathways of the corresponding viral proteins. A pairwise distance matrix (568x568) using the Jaccard distance (intersection over the union) was computed and used as input for clustering using R and the heatmap.2 function within gplots package (distance metric: euclidean, method: complete) (R Development Core Team, 2016; Warnes et al., 2016). Clusters were delimited after cutting the dendrogram at a specific height. Biological pathways significantly overrepresented within clusters were identified using a hypergeometric clustering with Bonferroni correction using as background the relative abundance of each biological pathway within the set of 568 human viruses. Dendrogram was formatted for visualization purposes using iTOL (Letunic and Bork, 2016).

Analysis of ZIKA interactome

ZIKV interacting human orthologs

We assembled a set of 690 mouse genes whose expression significantly changes in the brain of embryonic mice upon ZIKV infection (Li et al., 2016; Wu et al., 2016). Using this mouse gene set, we identified a set of 502 human ortholog genes, considered in our human dataset, with OrthoRetriever and the Ensembl BioMart version 77 (Kasprzyk, 2011). Significance of the overlap between the set of 97 human proteins predicted to interact with ZIKV proteins at an $LR_{P\text{-HIPSTER}} \geq 100$ and the set of 502 human orthologs was carried out with a hypergeometric test.

Protein interaction network analysis

In order to build the Protein Interaction Network (PIN) between ZIKV and human proteins we considered only the predicted viral-human PPIs with an $LR \geq 100$. We retrieved known protein interactions between the human proteins predicted to interact with ZIKV using the Intact database (Kerrien et al., 2012). The final interaction network combines both predicted viral-human PPIs and known human PPIs. Visualization of the network was carried out with cytoscape (Shannon et al., 2003). We clustered human proteins within the network based solely on their connectivity. First, we applied the MCODE plugin (Bader and Hogue, 2003) to find clusters of densely interconnected human proteins denoting potential functional modules or parts of pathways (include loops: no; degree cutoff: 2; haircut: no; fluff: no; node score: 0.4; kcore: 2; max depth: 100). Second, non-clustered proteins were grouped together according to their connectivity with ZIKV proteins. Enrichment analysis for each cluster was carried out with geneAnalytics (Ben-Ari Fuchs et al., 2016), only the five most enriched GO biological processes and phenotype sets were considered (minimum FDR corrected P value ≤ 0.0001). Bar plots were drawn with GraphPad Prism (Software).

Cells and virus

293T cells were maintained in Dulbecco's modified Eagle's medium (DMEM; GIBCO, Life Technologies) supplemented with 10% fetal bovine serum (FBS; Hyclone Laboratories) and 1% penicillin and streptomycin (GIBCO, Life Technologies) at 37°C and 5% CO₂. Zika virus (ZIKV), MR 766 strain obtained from BEI Resources was kindly provided by Dr. Vincent Racaniello (Columbia University Medical Center), was amplified once on Vero cells. Virus titer was determined by focus-forming assay on Vero cells.

Focus-forming assay

Vero cells were seeded into 24-well tissue culture plates at concentration of 80,000 cells/well. Serial 10-fold dilutions of each sample were prepared and added (in duplicates) to cell monolayers. Following 1hr incubation at 37°C, a semi-solid overlay containing 0.8% methylcellulose (Sigma-Aldrich), 3% fetal bovine serum, 1% Penicillin-Streptomycin in DMEM was added and plates were incubated at 37°C and 5% CO₂ for 48 hr. The semisolid overlay was then removed, cells were washed 3 times with PBS, and fixed with an acetone and methanol (1:1) solution for 30 min at -20°C. Cells were then subjected to immunohistochemical staining with mouse anti-flavivirus D1-4G2-4-15 antibody (EMD Millipore), incubated overnight at room temperature, followed by mouse IgG HRP-conjugated antibody (R&D Systems) for 1hr. This was followed by incubation with vector VIP peroxidase substrate (Vector Laboratories) until color developed. The number of foci was determined and used to calculate virus titers expressed as FFU/ml.

Gain of function

293T cells were seeded on to 12-well tissue culture plates at concentration of 150,000 cells per well for 24hr and transfected with 0.5 μ g of plasmids encoding indicated proteins (ORFeome collection 8.1) as indicated using TransIT-LT1 Reagent (Mirus) according to manufacturer's instructions. IFN β (PBL Interferon Source) priming was achieved by pretreatment at concentration of 1U/mL for 18hr prior virus infection. Cells were infected with ZIKV at multiplicity of infection (MOI) of 1 40hr post-transfection. Following a 1hr incubation at 37°C, cell monolayers were washed to remove unbound virus and incubated at 37°C and 5% CO₂ for 48hr. The cell supernatants was then harvested and virus was quantitated by focus-forming assay.

siRNA Transfection

Human breast adenocarcinoma cells (MCF-7) were seeded into 48-well tissue culture plates at concentration of 25,000 cells per well and transfected (in quadruplicates) with 50nM (final concentration) siRNAs duplexes using Lipofectamine RNAiMAX Transfection Reagent (Invitrogen, Life Technologies) according the manufacturer's instructions. The following siRNA were used: human ESR1 (SMARTpool, Dharmacon); AllStars Negative Control (QIAGEN) and AllStars Hs Cell Death (QIAGEN). 72hr post-transfection, cells were incubated with ZIKV (MOI of 1) for 1hr at 37°C. Cell monolayers were then washed to remove unbound virus and incubated at 37°C and 5% CO₂ for 48hr. Cell supernatants were harvested for virus quantification by focus-forming assay and cells were lysed with QIAzol reagent (QIAGEN) for real-time quantitative PCR (qPCR) analysis.

Real-time quantitative PCR

Cell samples were resuspended in QIAzol (QIAGEN) and total RNA was extracted following manufacturing instructions. cDNA was synthesized by using High Capacity cDNA Reverse Transcription Kit (Applied Biosystems). Real-time quantitative PCR (qPCR) was performed by using iTaq Universal SYBR Green Supermix (Bio-Rad Laboratories) in the CFX96 real-time PCR system (Bio-Rad Laboratories). The sequences of the primer pairs were the following:

GAPDH: 5'-ACCACAGTCCATGCCATCAC-3' and 5'-TCCACCACCCTGTTGCTGTA-3';

ESR1: 5'- GGGAAAGTATGGCTATGGAATCTG-3' and 5'-TGGCTGGACACATATAGTCGTT-3'; ZIKV: 5'- CCGCTGCCCAACACA AG-3' and 5'- CCACTAACGTTCTTTTGCAGACAT-3'.

Gene expression was calculated based on C_t values by using the formula $2^{-[C_t(\text{target gene}) - C_t(\text{GAPDH})]}$.

Classification of Human Papillomaviruses (HPVs)

LR-HPVs and HR-HPVs in dataset

Classification of HPVs into low risk cervical cancer and high risk cervical cancer was obtained from the literature (Cubie, 2013; IARC, 2019; Muñoz et al., 2003). Using this classification, we identified HR-HPVs and LR-HPVs in our dataset. HR-HPVs: HPV-16, HPV-18, HPV-39, HPV-59, HPV-68; LR-HPVs: HPV-6, HPV43, HPV-54, HPV-72, HPV-81.

HPV-interacting human proteins (Groups I and II):

A human protein is considered to preferentially bind viral proteins in a particular HPV category (LR-HPV or HR-HPV) whenever it is predicted to interact with viral proteins in at least half of HPVs in one category and less than half of HPVs in the remaining category, providing that the difference between number of interacting viruses in each category is at least half of the average category size:

$$b_{c1} \geq n_{c1}/2 \quad (2)$$

$$b_{c2} < n_{c2}/2 \quad (3)$$

$$b_{c1} - b_{c2} \geq \frac{(n_{c1} + n_{c2})}{2} \quad (4)$$

where b_{c1} is the number of viruses in class A (e.g., HR-HPV) with viral proteins predicted to bind a human protein, n_{c1} is the number of viruses in class A, b_{c2} is the number of viruses in class B (e.g., LR-HPV) with viral proteins predicted to bind the same human proteins, n_{c2} is the number of viruses in class B. A human protein is considered to preferentially bind one HPV category whenever it meets all three criteria described above.

HPV-interacting human proteins (Group III):

In order to identify the set of human proteins predicted to preferentially bind HPV viral proteins, we assessed the significance of the interaction with HR-HPVs and LR-HPVs viral proteins with a hypergeometric test using as background the entire predicted viral-human interactome. Only those human proteins predicted to interact with viral proteins in both HR-HPVs and LR-HPVs with a P value < 0.05 are considered as HPV-specific interacting proteins. Additionally, we included p53, which is predicted to interact with viral proteins from both LR and HR HPVs (3/5 and 4/5 respectively) highlighting the accepted role of this human protein in the life cycle of both LR- and HR HPVs (Li and Coffino, 1996; Pietsch and Murphy, 2008).

Bayesian network

A Bayesian network was trained using Weka (Frank et al., 2016). Training was carried out using the five HR-HPVs and five LR-HPVs. Each virus is described by the set of predicted interactions with the 10 human proteins, that were previously identified (see above),

using a binary code (1 if the human protein is predicted to interact with at least one of the corresponding viral proteins in a virus with an $LR_{P\text{-HIPSTER}} \geq 100$, 0 otherwise). Evaluation of the network was carried out by five-fold and leave-one-out cross-validation. Additionally, we also trained a Bayesian network classifier, where we applied an automatic and supervised feature selection method prior to training in order to ensure that the considered features highly correlate with the class (LR-HPV or HR-HPV) while having a low inter-correlation between them (Hall, 1999). Automatic selection of features reported the interactions with 8 human proteins, which were previously identified as proteins preferentially binding either HR-HPVs or LR-HPVs with the method described above. The additional two features (human proteins predicted to preferentially bind either HR- or LR-HPVs) not considered by the automatic supervised feature selection method (HNRNPM and TERF1) were redundant and did not contribute further to the performance of the classifier (which does not mean that the interaction with these human proteins are not biologically meaningful). As expected, five-fold and leave-one-out cross-validation of the Bayesian network trained considering only the interactions to eight human proteins reported the same accuracy (9 out of 10 viruses correctly classified as LR- or HR-HPV).

Clustering of HPVs

Each alpha-papillomavirus is described as a binary vector with 10 elements, each describing the predicted interaction with one of the 10 human proteins predicted to preferentially interact with viral proteins in HR-HPVs or LR-HPVs. Predicted interactions are converted into a binary vector using an LR cutoff of 100 (1 if the human protein is predicted to interact with at least one of the corresponding viral proteins in a virus with an $LR_{P\text{-HIPSTER}} \geq 100$, 0 otherwise). Clustering was carried out in R using the heatmap.2 package (distance metric: Manhattan; clustering method: complete) (R Development Core Team, 2016; Warnes et al., 2016).

Evolutionary analysis on human genes

Dataset

Primate datasets of 15,332 human coding genes were downloaded from the Ensembl genome browser (<https://useast.ensembl.org/index.html>) between August and November, 2017. The primate species included in the analyses were *Gorilla gorilla*, *Homo sapiens*, *Pan troglodytes*, *Pongo abelli*, *Nomascus leucogenys*, *Macaca mulatta*, *Papio Anubis*, *Chlorocebus sabaeus*, *Callithrix jacchus*, *Carlito syrichta*, *Microcebus murinus* and *Otolemur garnettii*. Only one sequence of each species was included. All datasets are publicly available from GITHUB (<https://github.com/RabadanLab/pamler>). Each nucleotide dataset was aligned as amino acids using Prank (Löytynoja, 2014), in order to preserve the correct frame, and then reconverted to nucleotides. The resulting alignment was then refined, by trimming regions of poor homology along the alignment with TrimAl (“gappyout” settings) (Capella-Gutiérrez et al., 2009).

Evolutionary analyses

A phylogenetic tree depicting the evolutionary relationships of the aforementioned primate species was built, using a topology reported previously in the literature (Hallast et al., 2016; Rogers and Gibbs, 2014). Given that sequences from all these 12 species were not necessarily present in a given gene, the original tree was pruned, thus retaining only the taxa for which there was a sequence available. Then, for each gene, the branch lengths of the primate tree were inferred with PhyML (Guindon and Gascuel, 2003) under the GTR + GAMMA (4 CAT) model, retaining the original tree topology.

The different genes analyzed were classified as “interacting” or “not-interacting,” according to the LR values inferred (using an $LR_{P\text{HIPSTER}}$ of 100). In order to assess the selective constraints in human proteins associated with the interaction with viral proteins, we performed site model analyses with codeml as implemented in PAML (Yang, 2007). Under the codon frequency model F3x4 (which estimates codon frequencies empirically), two different evolutionary models were compared by means of likelihood ratio tests (2 degrees of freedom): M7 (selective pressure of sites distributed according to a beta distribution, without positive selection) and M8 (selective pressure distributed according to a beta distribution, with the presence of positively selected sites). Significant evidence of selection was considered if $p < 0.05$, after False Discovery Rate correction. This model also allowed us infer the mean $\Delta N/\Delta S$ of each gene. $\Delta N/\Delta S$ values of Interacting (3,373) and not-interacting (11,601) genes were compared by means of Wilcoxon tests (2-tail).

Functional enrichment analyses

Enrichment analysis on human proteins with $\Delta N/\Delta S > 1$ and $\Delta N/\Delta S < 1$ (5th percentile) was carried out with DAVID (Huang et al., 2009a, 2009b). Enriched pathways were selected based on the nominal p value (< 0.05) and the False Discovery Rate correction (FDR $< 20\%$). For each $\Delta N/\Delta S$ subset, enrichments were carried out separately on: i) proteins predicted to interact with viral proteins with an $LR_{P\text{HIPSTER}} \geq 100$ and; ii) proteins not predicted to interact with viral proteins with viral proteins with an $LR_{P\text{HIPSTER}} \geq 100$.

Co-immunoprecipitation

293T cells were maintained in Dulbecco’s modified Eagle’s medium (DMEM; GIBCO, Life Technologies) supplemented with 10% fetal bovine serum (FBS; Hyclone Laboratories) and 1% penicillin and streptomycin (GIBCO, Life Technologies) at 37°C and 5% CO₂. Plasmids expressing the human proteins AMBP, BIRC2, BRCA1, DOCK5, ESR1, ING2, ISG15, RABGGTA, SFN, SPINT2, STAC2, STAT3, STK32A, TP53, VARS, VAV1, VCAM1, SPINLW1 (EPPIN) (ORFeome collection 8.1, Table S8), ABL1, APP, BIRC3, CUL1 (Table S8) and plasmid expressing the virus proteins HCV NS5A, HSV-1 UL13, HIV-1 Nef, ZIKV C, NS3 and NS5 (Addgene, Table S8) were used in this study. For co-immunoprecipitation (Co-IP), cells were seeded into 12-well tissue culture plates at concentration of 300,000 cells per well for 24 hr and cotransfected with 0.5 μg of each plasmid expressing the human and virus proteins using TransIT-LT1 Reagent (Mirus) (Figure S2). Single transfection with each plasmid were also performed to access individual protein expression (Figure S2). Cells from each well were lysed 36 to 40 hr after transfection in 0.2 mL of RIPA buffer (50 mM

Tris-HCl [pH 8.0]; 150 mM NaCl; 0.1% SDS; 1% Triton X-100; 0.5% Na Deoxycholate and EDTA-free protease inhibitor) at 4°C for 2 hr with constant agitation. Cell lysates were then sonicated for 2 min and centrifuged at 15,000 x g for 30 min at 4°C and pre-cleared for 1 hr at 4°C with protein G magnetic beads (SureBeads, Bio-Rad). The supernatant was then immunoprecipitated with 0.5 µg of monoclonal anti-HA antibody (Sigma-Aldrich) or 1:50 dilution of rabbit anti-FLAG antibody (Cell Signaling Technology) and 30 µL of protein G magnetic beads at 4°C overnight (O/N). The immunocomplex was then washed five times with 0.5 mL of RIPA buffer. Proteins bound to the beads were resuspended in 20 µL of Laemmli sample buffer (Bio-Rad) containing 5% β-mercaptoethanol (Sigma-Aldrich), boiled for 5 min at 95°C and then subjected to NuPAGE 4%–12% Bis-Tris Gel (Life Technologies) and transferred to 0.45 µm nitrocellulose membranes (Bio-Rad). Membranes were blocked with TBST containing 5% non-fat milk at room temperature (RT) for 1 hr and incubated O/N at 4°C with the anti-V5-HRP antibody (Life Technologies) or c-Myc polyclonal antibody (Santa Cruz Biotechnology) followed by the HRP-conjugated secondary antibody (Thermo Scientific) for 2 hr at RT. For detection, membranes were incubated with SuperSignal West Femto Maximum Sensitivity Substrate (Thermo Scientific) and imaging using ChemDoc Touch Imaging System (Bio-Rad).

The following lines contains a description of the used plasmids:

- i) The plasmid pDONR223-ABL1 was a gift from William Hahn & David Root, Addgene plasmid # 23939 (Johannessen et al., 2010). This plasmid was cloned into pLX304 Destination vector using LR Recombination Reaction (Gateway Technology, Invitrogen) according manufacturer protocol.
- ii) The APP gene was PCR amplified (attB1-APP_FOR 5'GGGGACAAGTTTGTACAAAAAAGCAGGCTTCACCATGCTGCCCGGTTTGGCACTGCTCCT3' and attB2-APP_REV 5'GGGGACCACCTTTGTACAAGAAAGCTGGGTTCCAGCATGAGCCATCGTGCCTGGCC3') from the human ORFeome Collection 8.1 plasmid (ORF ID 8905) and the purified PCR product was cloned into the pDONR223 vector using BP Recombination Reaction (Gateway Technology, Invitrogen) and further cloned into pLX304 Destination vector using LR Recombination Reaction (Gateway Technology, Invitrogen) according manufacturer protocol.
- iii) The plasmid Flag-clAP2/pRK5 was a gift from Xiaolu Yang, Addgene plasmid # 27973 (Hu et al., 2006). clAP2 is a ubiquitin protein ligase for BCL10 and is dysregulated in mucosa-associated lymphoid tissue lymphomas. The tag from this plasmid was replaced with a 3xMyc tag generated by an assembly PCR protocol (<https://primerize.stanford.edu>). The following oligos were used to generate the 3xMyc tag containing EcoRI and XbaI restriction sites at the 5' and 3', respectively: PAC 1 For 5' TCGATTGAATTGCGCGCCATGGAGCAGAACTCATCTCTGAAGAAGATCTGGAACAAA 3'; PAC 2 Rev 5' CCAGATCTTCTCTGAAATCAACTTTTGTCCAGATCTTCTTCAGAGATGAGTTTCTGC 3'; PAC 3 For 5' AGTTGATTTCCAGAAGAAGATCTGGAACAGAAGCTCATCTCTGAGGAAGATCTGGG 3'; and PAC 4 Rev 5' TACTATGTTTCTAGAGATCCCAGATCTTCTCAGATGATGAGCTT 3'.
- iv) The plasmid pcDNA3-myc3-CUL1 was a gift from Yue Xiong (Addgene plasmid # 19896) (Ohta et al., 1999).
- v) The plasmid pCMV-Tag1-NS5A was a gift from Xin Wang (Addgene plasmid # 17646) (Budhu et al., 2007).
- vi) The plasmid HSV-1 UL13 was a gift from Robert Kalejta (Addgene plasmid # 26697) (Kuny et al., 2010).
- vii) The plasmid pCI NL4-3 Nef-HA-WT was a gift from Warner Greene (Addgene plasmid # 24162) (Geleziunas et al., 2001).
- viii) The plasmid pLV_Zika_Cv_Flag was a gift from Vaithi Arumugaswami (Addgene plasmid # 79628): Unpublished.
- ix) The plasmid pLV_Zika_NS3_Flag was a gift from Vaithi Arumugaswami (Addgene plasmid # 79635): Unpublished.
- x) The plasmid pLV_Zika_NS5_Flag was a gift from Vaithi Arumugaswami (Addgene plasmid # 79639): Unpublished.

Webserver development

The P-HIPSTer (<http://phipster.org>) webserver base code is split into two distinct parts: a back-end, which comprises a rest API written in the Django framework; and a front-end, which provides the user-interface to query this API, process the structured data it returns, and display the results. The API layer communicates with a Postgres database. The front-end is coded in the javascript framework Vue.js.; and was written as a single-page-application (SPA). It features the third party module NGL Viewer for molecular visualization (Rose et al., 2018). We host the webserver on Amazon Web Services (AWS).

In addition to the set of predicted PPIs, the P-HIPSTer webserver provides, for each viral protein, sequence-, structure- and PPI-based functional annotation, the corresponding Pfam (Finn et al., 2016) domains, a graphical representation of the predicted protein interaction network, and a molecular viewer that allows to explore the structural models of viral and human proteins as well as the predicted interaction complexes, whenever possible (the LR_{dom} must be ≥ 100).

Sequence-based annotation

We identify the set of homologs after Blasting the sequence of a viral protein against the Uniprot database, considering as homologs only those hits with maximum E-value of 10^{-20} , a minimum alignment coverage of 70% of the shortest sequence aligned and a minimum sequence identity of 90% (Altschul et al., 1990; Apweiler et al., 2004). The webserver reports the union of GO terms (Ashburner et al., 2000; The Gene Ontology Consortium, 2017) extracted from each homolog, ranked based on their frequency.

Structure-based annotation

A structural neighbor search is carried out for each viral protein or domain (either taken directly from the PDB or modeled by homology) against the PDB database with Ska (Petrey et al., 2003; Yang and Honig, 2000b) using a Protein Structural Distance (PSD) ≤ 0.6

and a Structural Alignment Score (SAS) ≤ 3.5 Å. The webserver reports the union of GO (Ashburner et al., 2000; The Gene Ontology Consortium, 2017) and EC (Bairoch, 2000) terms extracted from each structural neighbor, ranked based on their frequency.

PPI-based annotation

See Functional enrichment analysis using Gene Set Enrichment Analysis (GSEA).

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical parameters, including the definition of center, dispersion and associated significance, are reported in the main text, Figures, Figure legends and Tables. We have applied hypergeometric test, permutation test and Wilcoxon test to calculate significance. Whenever appropriate, p values were adjusted for multiple comparisons. The section entitled “[Method Details](#)” describes the statistical analyses performed. Data are judged to be statistically significant when $p < 0.05$ in applied statistical analyses. In [Figure 3](#), asterisks denote statistical significant (*, $p < 0.05$; **, $p < 0.0001$).

DATA AND CODE AVAILABILITY

All data generated as part of this study is available at <http://pHipster.org> and GITHUB (<https://github.com/RabadanLab/pamler>). In addition, P-HIPSTer code is made available upon request.

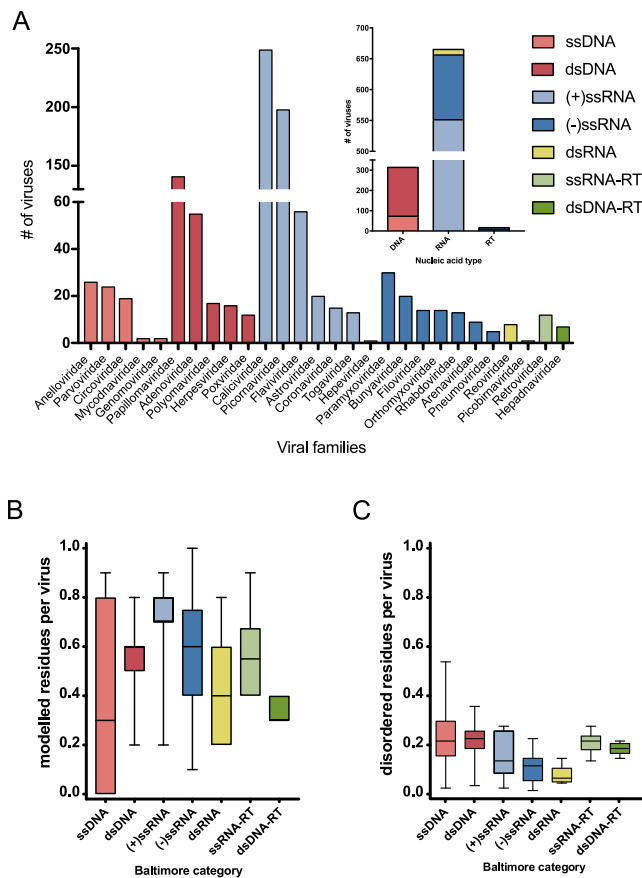


Figure S1. Summary of Viral Families and Baltimore Categories Interrogated by P-HIPSTER and Homology Modeling Coverage across Baltimore Categories, Related to Figure 1

a) Fully sequenced viruses infecting humans were downloaded from VirusHostDB (<https://www.genome.jp/virushostdb/>). For a complete list of viruses and Baltimore category as well as Family membership, see Table S1. b) Fraction of modeled residues per virus within each Baltimore category; c) Fraction of disordered residues per virus within each Baltimore category. Structure modeling was performed as described in methods. Center lines correspond to median values and whiskers range from min to max.

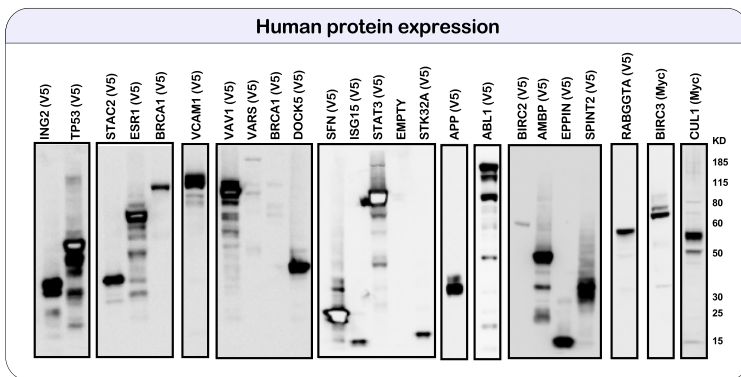
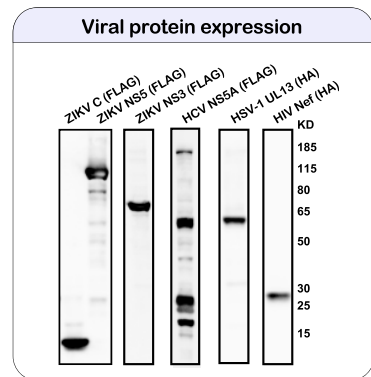
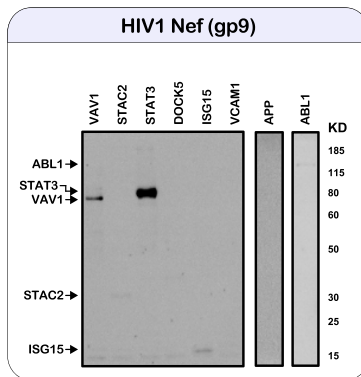
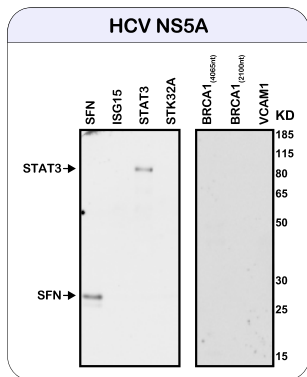
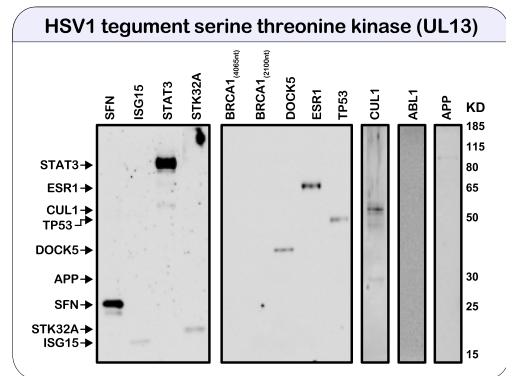
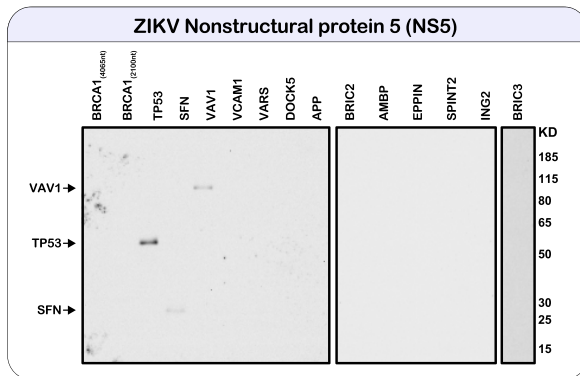
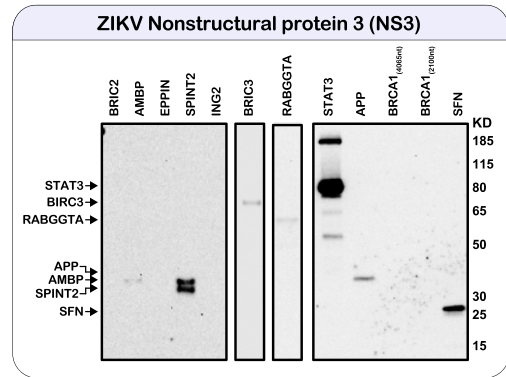
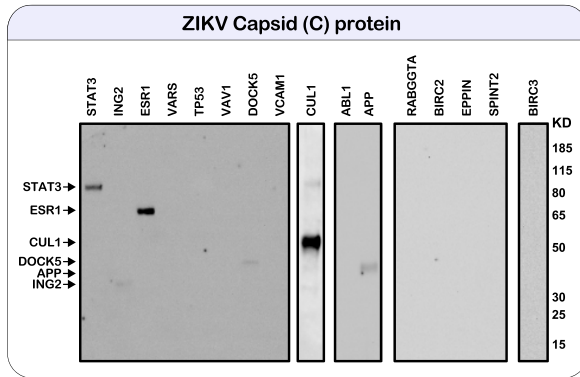


Figure S2. Empirical Validation of P-HIPSTer Predictions, Related to Figures 1 and 2

65 P-HIPSTer predictions spanning LR values between 1.2 and 1,106 were selected for co-IP. Briefly, 293T cells were transfected with relevant expression plasmids and lysates were subjected to immunoprecipitation with indicated antibodies followed by SDS-PAGE electrophoresis and immunoblotting. Shown are representative gels. All experiments were performed at least 3 times to ensure reproducibility. For further details, please refer to [STAR Methods](#). A complete summary of validation experiments is available in [Table S3](#).

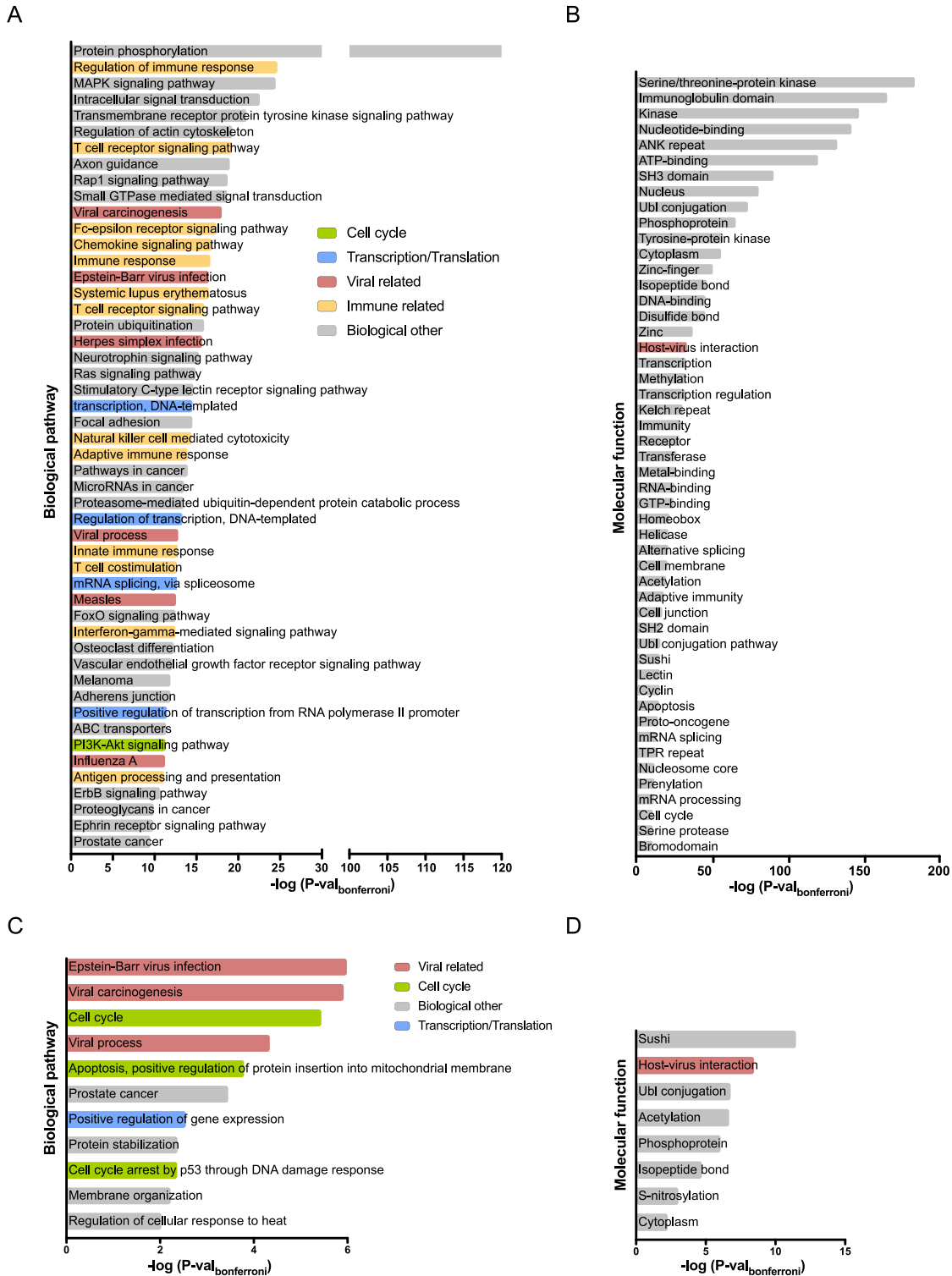


Figure S3. P-HIPSTER Uncovers Viral Targeting of Infection-Related Pathways, Related to Figures 1 and 5

Enriched biological pathways (a) and molecular functions (b) within a set of 5,749 human proteins predicted to interact with human viruses. Enriched biological pathways (c) and molecular functions (d) within a set of 173 human proteins predicted to interact with ≥ 100 human viruses.

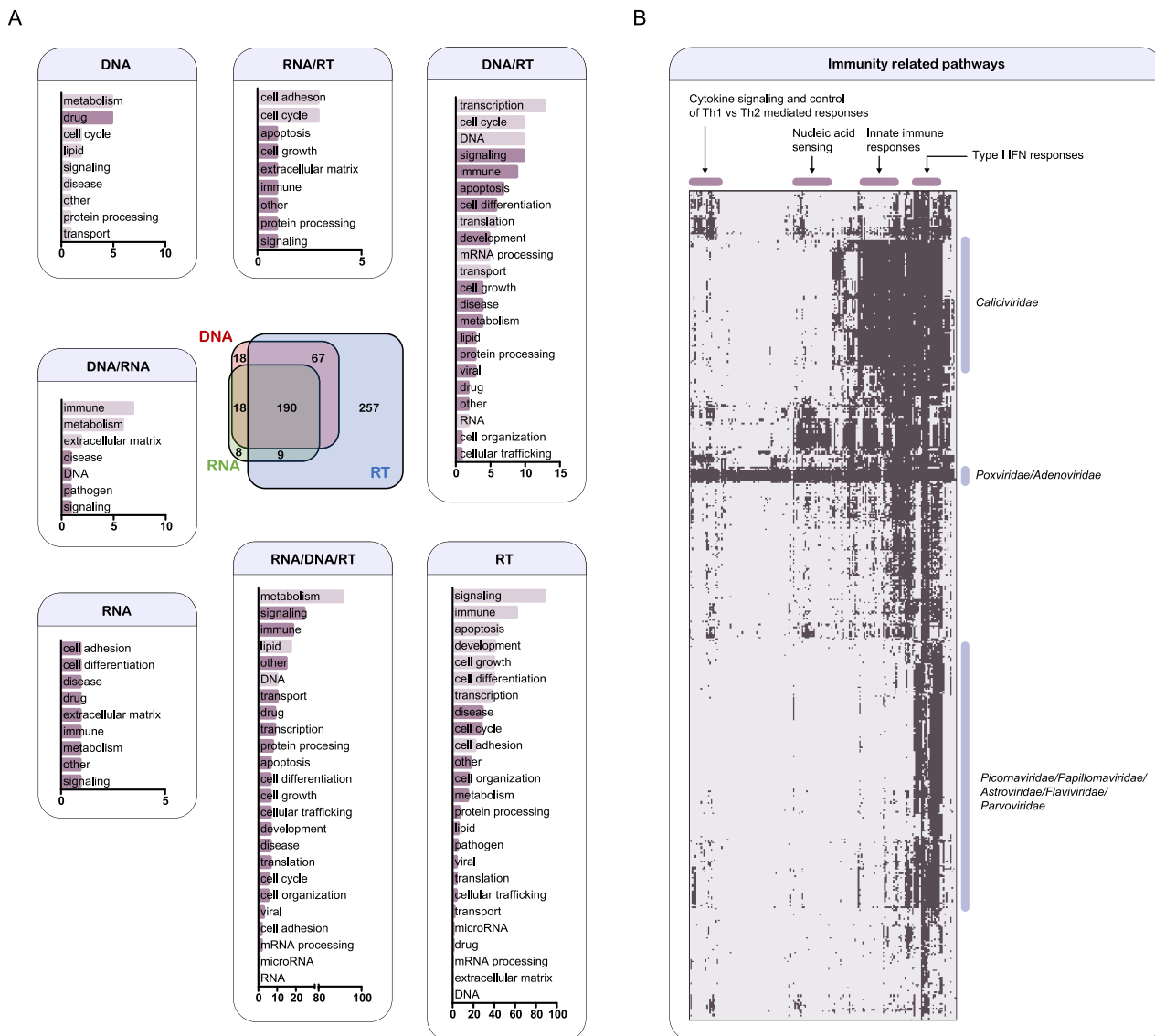


Figure S4. P-HIPSTER Uncovers Shared and Unique Machinery Employed across Viruses of Differing Nucleic Acid Type, Related to Figure 5

a) Venn diagram illustrating overlap of enriched biological pathways within each nucleic acid type (indicated are number of pathways within each region of the Venn diagram). Shown are pathway categories that are over-represented (x axis denotes number of enriched pathways within each category). Categories that are significantly enriched within each region of the Venn diagram are highlighted in light purple. b) Clustering of immunity-related pathways and viruses based on predicted pathway enrichment. Dark cells within the heatmap correspond to pathways enriched with q-value < 0.01. Highlighted are pathway classes and viral families corresponding to co-clustered pathways and viruses.

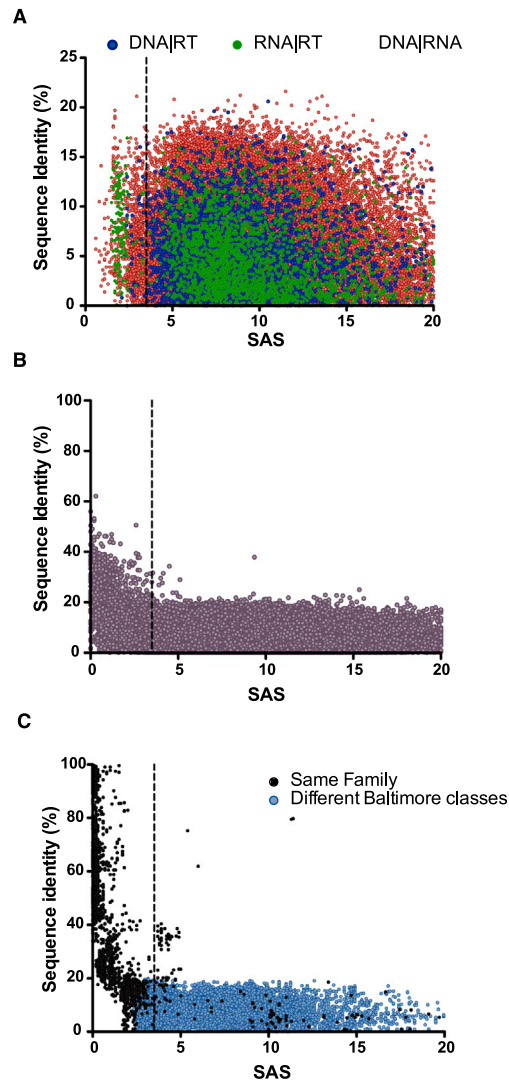


Figure S5. Comparison of Structural and Sequence Similarity across Viral Protein Pairs, Related to Figure 5

a) Subset of viral proteins with low pairwise sequence identity (< 40%) where proteins in each pair corresponds to a virus from a different nucleic acid type: DNA/RNA (Red), DNA/RT (Blue), RNA/RT (Green). b) Subset of viral proteins across all nucleic acid types with low pairwise sequence identity (< 40%). c) Viral capsid proteins corresponding to viruses from the same viral family (Black) or different Baltimore categories (Blue). Vertical bars indicate threshold to consider a pair of viral proteins as global structural neighbors (Structural Alignment Score $SAS \leq 3.5A$).