1. Title page

Title: Consensus clustering of herpesvirus protein interaction networks provides insights into their evolutionary relationship with bacteriophages

Subtitle: New insights on herpesviruses evolution from biological network analysis

Anna Hernández Durán[1,2], Todd M. Greco[3], Benjamin Vollmer[2], Ileana M. Cristea[3], Kay Grünewald[2,4†], Maya Topf[1,†]

[1] Institute of Structural and Molecular Biology, Birkbeck College, University of London, Malet street, London, WC1E 7HX, UK;

[2] Division of Structural Biology, Wellcome Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford, OX3 7BN, UK;

[3] Department of Molecular Biology, Princeton University, Lewis Thomas Laboratory, Washington Road, Princeton, NJ 08544, USA;

[4] Department of Structural Cell Biology of Viruses, Centre for Structural Systems Biology, Heinrich Pette Institute, Leibnitz Institute of Experimental Virology, University of Hamburg, Notkestr. 85, Hamburg, D-22607, Germany

[†]To whom correspondence should be addressed: E-mails: m.topf@cryst.bbk.ac.uk; Tel.: +44 (0)20 7631 6886; E-mail: kay.gruenewald@cssb-hamburg.de, Tel. +49 (0)40 8998 87700.

Abbreviations

CAM: consensus agreement matrix

CLR: collagen-like repeat

DIP: Database of Interacting Proteins

dsDNA: double-stranded DNA

EBV: Epstein-Barr virus

EGFP: enhanced green fluorescent protein

EMDB: Electron Microscopy Data Bank

ePPI: experimentally detected PPI

GO: gene ontology

HCMV: human cytomegalovirus

HMM: Hidden Markov Model

HPI: hours post infection

HSV1: herpes simplex virus type 1

IP: immunoaffinity purification

LCC: largest connected component

MCP: major capsid protein

MS: mass spectrometry

PDB: Protein Data Bank

PPI: protein-protein interaction

oPPI: PPI in orthologous species

pPPI: computationally predicted PPI

TGN: trans-Golgi network

TMT: tandem mass tags

VZV: varicella-zoster virus

2. Abstract

Infections with human herpesviruses are ubiquitous and a public health concern worldwide. Current treatments reduce the morbidity of some manifested symptoms but neither remove the viral reservoir from the infected host nor protect from the recurrent symptom outbreaks that characterize herpetic infections. The difficulty in therapeutically tackling these viral systems stems in part from their remarkably large proteomes and the complex networks of physical and functional associations that they tailor. This study presents our efforts to unravel the complexity of the interactome of herpes simplex virus type 1 (HSV1), the prototypical herpesvirus species. Inspired by our previous work, we present an improved computational pipeline for the protein-protein interaction (PPI) network reconstruction in HSV1, combining both experimentally supported and bioinformatic predictions. Our newly-developed consensus clustering approach allows us to extend the analysis beyond binary physical interactions and reveals higher order functional associations including that of pUS10 with capsid proteins. In-depth bioinformatics sequence analysis unraveled structural features of this 34-36 kDa protein reminiscent of those observed in some capsid-associated proteins in tailed bacteriophages, with which herpesviruses are thought to share a common ancestry. This suggests that pUS10 could represent an evolutionary vestige between these two viral lineages. Using immunoaffinity purification-mass spectrometry we found that pUS10 specifically co-isolated with the inner tegument protein pUL37, which binds cytosolic capsids,

contributing to initial tegumentation and eventual virion maturation. In summary, this study unveils new insights at both the system and molecular levels that can help better understand the complexity behind herpesvirus infections.

3. Introduction

Herpesviruses infect a wide range of eukaryotic organisms, and are the etiologic agent of severe diseases in livestock and humans. At present, there are nine species of herpesviruses known to routinely infect human. They are referred to as the human herpesviruses, and their infections are associated with symptoms ranging from fever and cutaneous lesions, to encephalitis, meningitis, and a number of cancerous malignancies [1]. A link between herpetic infections and the neurodegenerative Alzheimer's disease was recently confirmed, emphasising the socio-economic burden associated with these viruses [2,3].

Herpesviruses are enveloped viruses that assemble into a morphologically unique extracellular particle (i.e. virion), which is organised in concentrical structural layers [4]. At the innermost of the virion particle, an icosahedral protein shell of ~ 120nm diameter - the capsid - encloses the double-stranded DNA (dsDNA) genome of the virus. Surrounding the capsid, a protein matrix - the tegument - occupies about two thirds of the virion particle volume. The tegument contains proteins that are delivered into the host cytoplasm upon infection and therefore are ready to initiate their function prior to the transcription of any viral genes. The entire particle is coated by a lipid bilayer -

the envelope - that contains several proteins and glycoproteins that are crucial for host cell entry and cell-to-cell spread, as well as for modulation of the host's immune response. As for other enveloped viruses, entry into the host cell occurs via fusion of the envelope with plasma membrane or endocytic cellular membranes, which results in release of the virion content into the host cell cytosol [5]. Altogether, herpesviruses are composed of between ~70 to ~170 different protein species and achieve diameters between ~150 to ~250nm [6,7].

To understand the behavior of a system, the relationships among the components are as important as the components themselves. The need for a comprehensive description of the relationships among viral proteins, and with host factors, has prompted numerous protein-protein interaction (PPI) studies. In 2016 we published a compilation of PPI data in herpes simplex virus type 1 (HSV1), the prototypical species of the *Herpesviridae* family based on integration of experimentally supported data from five public resources [8]. Here, we present an improved computational pipeline for PPI network reconstruction, collating data from seven public repositories with computationally predicted PPIs inferred using conservative sequence similarity assessments. The reliability of all gathered PPIs is measured using a new scoring scheme. Moreover, we applied network analysis approaches to gain insights into the functional organization of the extracellular viral particle. We developed a clustering consensus protocol to study the community structure underlying the binary interactions among virion proteins. The resulting higher order relationships predicted previously unrecognized

functional viral protein associations, such as between the inner tegument protein pUL37 and pUS10. We further investigated this relationship using bioinformatic sequence analysis, and provided experimental support using immunoaffinity purification-mass spectrometry in primary human fibroblasts infected with HSV1.

4. Results

PPI network assembly

Binary Protein-Protein interaction (PPI) data were obtained from five molecular interaction repositories (BioGRID [9], the Database of Interacting Proteins (DIP) [10], IntAct [11], Mentha [12], and VirHostNet 2.0 [13]) and two structural databases (Protein Data Bank (PDB) [14] and the Electron Microscopy Data Bank (EMDB) [15]) (Fig 1A). We collected PPIs that have been detected for all nine human herpesvirus species, and three closely-related non-human herpesviruses (Fig 1B and S1 and S2 Tables). The latter were included as these species are frequently used as animal models of herpetic human infections (see Materials and Methods). The resulting joined non-redundant data set contained 2854 unique pieces of evidence, 2363 unique PPIs, and 758 unique protein sequences. From this data set, PPIs experimentally detected in HSV1 were directly added to the network (Fig 1C). PPIs experimentally detected in species other than HSV1 were used to predict new PPIs in HSV1 (Fig 1D). To predict PPIs, we used an orthology-based method referred to as *interologues mapping* [16]. This method predicts an interaction in HSV1, if both putative interactors have homologues known to

interact in other species (see Materials and Methods). Homology relationships are inferred based on sequence-based Hidden Markov Model (HMM) profile alignments [17] in combination with a conservative multi-criteria threshold to filter out potential spurious matches from the homology search results (see Materials and Methods).

Fig 1. Network assembly framework. (A and B) PPI data for a total of twelve herpesvirus species (nine human and three non-human herpesviruses, together covering members of all three subfamilies, i.e. the α-, β- and γ-*Herpesvirinae,* S1 and S2 Tables) were collected from seven public resources [9-15]. (C) PPIs experimentally detected in HSV1 were transferred to its interactome (ePPIs). (D) PPIs detected in any of its orthologous herpesvirus species (oPPIs) were used to predict PPIs in HSV1 (pPPIs). Predictions were conducted based on a sequence-based interologues mapping [16] approach (green box), and included the following steps: for each protein involved in a binary oPPI, (1) sequence-based homologous sequences in the HSV1 proteome were searched for using HHblits [17]; (2) a conservative homology threshold was applied to filter out potential spurious matches among the list of candidates returned by HHblits. From the remaining matches, the best scoring sequence was selected as the most reliable putative HSV1 homologue; (3) if potential HSV1 homologous sequences were found for both proteins in the initial oPPI, an interaction between the two HSV1 sequences was predicted. (E) Predicted and experimentally supported PPIs were joined into a non-redundant data set, and scored based on their supporting evidence (see Materials and Methods).

Experimentally detected interactions in HSV1 and computational predictions were compiled into a PPI network, and the confidence of each interaction based on its cumulative evidence was assessed using a new scoring function (see Materials and Methods). The new function is a modification of the MIscore function [18], which was developed by the Proteomics Standards Initiative [19], and is compliant with standardised protocols for the assessment and representation of molecular interaction data [20]. Our function essentially adds a conservative penalty function term to the original MIscore function for interactions without experimental support. This term includes information on the prediction method (in this case sequence-based alignment) and the total number of homologous species from which the prediction was inferred. The function gives higher confidence scores to interactions that have been predicted from a larger number of species. For instance, if an interaction can be predicted in HSV1 based on interactions experimentally detected in three other species of herpesviruses (e.g. VZV, EBV and HCMV), the prediction will score higher than if it had been obtained from a single experimental observation (e.g. only in VZV).

The reconstructed network (Fig 2 and S3 Table) contained 369 PPIs, formed among 68 proteins, and supported by 643 unique pieces of evidence. Among the 369 PPIs, there were 159 experimentally supported interactions, and 250 computationally predicted; 40 interactions were supported by both experimental and computational evidence. All these network data are

200 available through our new version of the HVint database [8]: http://topf-
201 group.ismb.lon.ac.uk/hvint/.
202
203
204 **Fig 2. Reconstructed PPI network for HSV1**. Nodes represent proteins and
205 their size is proportional to the number of interactions associated with each
206 protein in the network (degree). Nodes are colour-coded as follows: cyan for
207 capsid and capsid-associated proteins, orange for tegument proteins, yellow
208 for envelope glycoproteins, blue for non-glycosylated envelope proteins,
209 respectively, and grey for proteins that are not present in the mature virion
210 particle (i.e. typically only expressed during intra-cellular stages). Edge (or
211 link) thickness reflects the confidence score for the interaction (the thicker, the
212 higher the confidence). Edges are colour-coded to indicate the type of
213 supporting evidence behind it, i.e. blue for experimentally supported
214 interactions, red for computationally predicted, and green for interactions with
215 both experimental and computational supporting evidence.
216
217
218 **Selection of high confidence PPI predictions for experimental**
219 **testing**
220 We assessed the ability of our protocol for predicting meaningful interactions
221 for experimental testing. We conducted a manual literature review in search of
222 experimental support for our best predictions (5% top scoring interactions)
223 among published literature that had not been included in the input dataset (i.e.
224 it was not used to infer the predictions). This search returned experimental

support for 4 out of the total 11 (top 5% scoring) predicted PPIs examined [8,21,22] (Table 1).

Table 1. Top 5% computationally predicted PPIs.

Interacting protein A	Interacting protein B	Score
UL32 (P10216)	UL32 (P10216)	0.347
UL32 (P10216)	UL40 (P10224)	0.297
UL5 (P10189)	UL52 (P10236)	0.287
UL25 (P10209)	UL25 (P10209)	0.266
UL15 (P04295)	UL50 (P10234)	0.263
UL40 (P10224)	UL52 (P10236)	0.26
UL39 (P08543)	UL50 (P10234)	0.257
UL23 (P03176)	UL33 (P10217)	0.254
UL21 (P10205)	UL33 (P10217)	0.254
UL5 (P10189)	UL16 (P10200)	0.249
UL12 (P04294)	UL16 (P10200)	0.236

Interacting proteins (A and B) are referred to with their open reading frame name and UniProtKB accession numbers. Shaded rows indicate predicted interactions with additional experimental support not included as input during network reconstruction.

Analysis of the community structure in the HSV1 network

Using the HSV1 network data compiled at this point, we undertook the analysis of the community structure of the subnetwork formed by virion proteins. These are defined as components of the extracellular viral particle (i.e. capsid, tegument, and envelope proteins), which are currently well identified in this species [23-26]. We applied two further steps to remove noise (e.g. false positive PPIs) in the network to reduce potential sources of artefacts in the clustering results (Fig 3A, see Materials and Methods). First, we removed PPIs in the network taking place between capsid (and capsid-associated) and envelope proteins, which are likely unfeasible due to spatial constraints imposed by the virion tegument. Second, we extracted, from the network obtained in the previous step, the largest connected component (LCC – the largest group of nodes in which any pair of them can be connected by a path). The resulting network contained 40 nodes and 94 edges.

Next, 14 different clustering algorithms were applied to this LCC network (Fig 3B and S4 Table). Those output partitions yielding positive modularity (11 in total) were integrated into a consensus agreement matrix (Fig 3C). Values in the consensus agreement matrix cells represented, for each pair of proteins in the network, the fraction of partitions that clustered the pair together. Using a threshold-based method guided by modularity maximization a consensus partition was inferred from the consensus agreement matrix (Fig 3D). This partition achieved a modularity of 0.43 (for thresholds in [0.55, 0.60]) (S1 Fig and S5 Table); values between 0.3 and 0.7 are considered indicative of a significant community structure [27,28].

Fig 3. Consensus clustering protocol. Starting from an input network, (A) a series of steps were applied to remove potential false positives. Here, these steps included selection of virion components only, and removal of PPI between capsid (and capsid-associated) proteins to envelope proteins. (B) N=14 clustering algorithms were applied to the network. (C) The output base partitions with positive modularity were integrated into a consensus agreement matrix. (D) A final consensus clustering partition was derived by filtering out the values in the consensus agreement matrix based on a modularity-maximisation threshold.

The consensus partition divided the network into four communities (Fig 4). Their biological consistency was then assessed using functional annotation data manually curated from Gene Ontology (GO) [29,30] annotations and published literature.

Fig 4. Community structure inferred for the HSV1 virion network. Depiction of nodes and edges follows the same criteria as described in Fig 2. Four separate communities are indicated by grey shaded areas.

Community A

Community A illustrates almost entirely the composition of the capsid (Fig 4). The only capsid protein missing is pUL6 (the portal protein), which appears disconnected from the LCC in the virion subnetwork. This was surprising as the localisation of pUL6 in the capsid, forming the capsid portal complex, is

well established [31]. However, the input data sets used to feed the network assembly framework did not contain evidence for this interaction.

Community A also contains the protein pUS10. pUS10 is a poorly characterised α-subfamily specific protein of 312 residues, which has been found in the nuclear, peri-nuclear and cytoplasmic cellular compartments and co-precipitating with capsids, yet direct interactions with capsid proteins have not been reported. It exists in two phosphorylation states and is regarded as a minor component of the tegument [32]. A consensus zinc finger was identified in pUS10 homologues although the protein failed to bind nucleic acids (common among zinc finger proteins) during experimental testing [32,33]. Finally, HSV1 pUS10 presents a high proline content at its N-terminus and a four-residue long polyproline sequence located centrally. To assess whether the clustering of pUS10 with capsid proteins had a functional significance or it was an artefact of a low number of interactions for pUS10 in the input graph, we sought further characterization of the protein through primary sequence analysis (Fig 5 and S2 Fig, see Materials and Methods). An initial search for potential sequence homologues did not identify any candidates beyond pUS10 counterparts, yet it highlighted the presence of seven tandem collagen-like repeats (CLRs) located toward the N-terminus of the protein sequence. Interestingly, through a manual curation of the literature, we found support for this prediction in a 30-year old study [34] (note that pUS10 was annotated by its MW in this study, which could explain why this prediction was not included in databases such as UniProtKB). This prominent feature prompted compelling hypothesis on its functions and evolutionary history (see

Discussion). Secondary structure predictions indicated that the N- and C-termini of the protein are structurally distinct. Whilst the N-terminus was predicted to be disordered, the C-terminus was rich in α-helices. Additionally, our predictions revealed a potential single-pass transmembrane segment at the very C-terminus of the protein. This prediction overlaps with the potential zinc finger motif, which could explain why so far no functional evidence has been provided.

Fig 5. Sequence characterization of pUS10. Both previously described and newly identified features are indicated. Predicted disordered regions, α-helices and transmembrane helices are highlighted in blue, yellow and pink, respectively. Collagen-like repeats (CLRs) are framed in red boxes. Prolines are highlighted in red. The four-residues polyproline sequence is framed in a black box. The previously found consensus zinc finger sequence [33] is underscored.

The binary interaction that connects pUS10 with other components within its community is the interaction with pUL37. This interaction was predicted by our orthology-based method, and we aimed to further validate it experimentally (see below). pUL37 HSV1 is an essential inner tegument protein [35] that contains a binding site for HSV1 pUL36, through which it is recruited to cytosolic capsids before secondary envelopment [36,37] or to sites of secondary envelopment in the absence of capsids [38].

Experimental support of the pUS10-UL37 interaction by immunoaffinity purification – mass spectrometry

To provide evidence in support of the predicted pUS10-UL37 interaction, we isolated pUL37 from infected cells using immunoaffinity purification and analysed the co-isolated proteins by quantitative mass spectrometry (IP-MS). Specifically, protein complexes were isolated from human fibroblasts synchronously infected with HSV1 strains encoding either pUL37 tagged with enhanced green fluorescent protein (pUL37-EGFP) or EGFP alone as a control. Isolations were performed at two functionally distinct time points of infection, 8 and 20 hours post infection (hpi) (S6 Table). Eight hpi represents an early time point of pUL37 expression, which is prior to secondary envelopment. This is consistent with our observation that pUL37-EGFP appears diffusely localized in the cytoplasm by epifluorescence microscopy (Fig 6A). In contrast, at 20 hpi secondary envelopment and virus particle release is in progress, reflected by pUL37-EGFP fluorescence visualized as capsid-associated puncta within the cytoplasm and maturing virions (Fig 6A). The observed temporal kinetics and localization of pUL37 induction were consistent with the prior characterization of this pUL37-EGFP HSV1 strain [37].

After establishing the temporal expression and cellular distribution of pUL37, we performed IP-MS analyses of pUL37-EGFP and EGFP complexes isolated from cytoplasmic-enriched lysates (Fig 6B). The isolation of pUL37-EGFP from the input lysates was confirmed by western blot (Fig 6C). Next, the co-

isolated proteins were identified and relatively quantified using labelling with isobaric with Tandem Mass Tags (TMT). Specifically, proteins were digested with trypsin, and the resulting peptides were derivatised with distinct TMT labels. This approach offered multiplexing, as the labelled pUL37-EGFP and EGFP IPs performed in biological replicates (N = 2 replicates) were combined and analysed by tandem MS (Fig 6B). Therefore, both the specificity of the pUL37 interactions (via comparison to the control EGFP IP) and the relative abundance of the interactions at the different time points of infection (8 and 20 hpi) could be simultaneously assessed in this quantitative MS workflow. Nine viral proteins, including pUS10, that appear as interacting with pUL37 in our network were quantitatively enriched by ≥2-fold in pUL37-EGFP vs EGFP control IPs in at least one-time point and in both replicate experiments (S6 Table). After normalizing the pUS10 abundance to the abundance of pUL37-EGFP bait calculated by TMT-MS, we observed that the relative amount of pUS10 co-isolated with pUL37 was increased at the later stage of infection. At this stage, pUL37-EGFP is localized to sites of secondary envelopment and associates with maturing virions, consistent with the formation of pUL37-EGFP puncta at 20 hpi (Fig 6A). These data suggest a potential role for the pUL37-US10 interaction in the cytoplasm during virion maturation.

Fig 6. Experimental validation of the pUL37-US10 interaction in HSV-1 infected primary human fibroblasts. (A) Visualization of pUL37-EGFP during HSV1 infection of human foreskin fibroblasts using live cell epifluorescence microscopy. Images show a representative field of infected cells at 8 and 20 hpi. Zoomed images show localization of pUL37-EGFP

(green) in the same cell at 8 and 20 hpi. Scale bar = 50 µm. (B) IP-MS workflow. Human fibroblasts were synchronously infected (multiplicity of infection = 10) with either pUL37-EGFP or EGFP HSV1, with two replicates per condition. HSV1-UL37GFP was collected at 8 and 20 hpi and HSV1-GFP at 20 hpi (HSV1-GFP). pUL37-EGFP and its interactions were isolated from the cytoplasmic cell fraction by immunoaffinity purification using anti-GFP antibodies. Proteins were digested with trypsin, and the resulting peptides from each sample were labelled with unique TMT reagents and then combined prior to nanoliquid chromatography-tandem mass spectrometry analysis. (C) The recovery of pUL37-EGFP and EGFP in the immunoisolates was assessed by western blot with an anti-GFP antibody. 10% of each sample was analysed. (D) PPIs around pUL37 present in the reconstructed HSV1 network (Fig 2) and supported by immunoaffinity purification results. (E) The abundance of pUS10 interaction with pUL37 at 8 and 20 hpi (average ± range, N=2). The relative amount of pUS10 was calculated by TMT-MS quantification and normalized by the respective pUL37 TMT abundance in each IP.

Community B

In community B (Fig 4), proteins pUL41, pUL47, pUL48, and pUL49 are associated, directly or indirectly, and are all known to modulate gene expression [39-41]. Another major group of proteins, (pUL11, pUL14, pUL16, pUL21, and pUL6) are related to early virion morphogenesis, i.e. recruitment of capsids and virion components to the enveloping membranes during early

407 stages of tegumentation [42-45]. The tight relationships among the gene
408 expression modulation proteins have been well characterized, yet the steps
409 that guide their incorporation into the virion particle are not yet clearly defined.
410 The results of our clustering analysis suggest that this process could be
411 guided by the coordinated action of cluster B members involved in virion
412 morphogenesis.
413
414 *Community C*
415
416 Community C is defined by proteins involved in late stages of secondary
417 envelopment and virion release. This community is enriched in envelope
418 glycoproteins (gM/gN, gK/UL20, pUL45, pUL43, gJ, and gG) that regulate
419 membrane-associated events. These include internalization of proteins from
420 the plasma membrane, membrane fusion rates, and modulation of immune
421 responses. Tegument proteins in this community are also annotated with the
422 trafficking of virion components at late stages of virion morphogenesis [46-52].
423
424 An interesting observation from community C comes from the presence of
425 pUL55, another poorly characterised α-subfamily-specific protein [53]. Our
426 primary sequence analysis of this 186-residue protein failed to find
427 characteristic structural features but predicted pUL55 to be an α/β protein.
428 The analysis of the hierarchical structure of community C on the other hand
429 (S3 Fig, see Materials and Methods), did suggest a very stable clustering
430 tendency with four other proteins in the cluster, namely gK/UL20, pUL45, and
431 pUS2. The cluster assignment together with the particularly strong association

with the latter four proteins [54-57] suggest new functional scenarios in late envelopment and virion release events for the currently poorly-understood pUL55 protein.

Community D

Community D is the most functionally inconclusive community. We hypothesise that this is due to the lack of known binding partners (either due to missing data in the input data sets or to failed homology mappings) in the network. Specifically, the lack of proteins gI and pUS9, both associated to gE (gI physically and pUS9 functionally [58,59], as well as proteins gB and gD, which are the members of the fusion machinery (together with gH/gL [5]).

5. Discussion

Human herpesviruses are ubiquitous human pathogens with a severe socioeconomic impact worldwide [1]. Key to a successful infection is the delicate coordination of the complex network of virus-virus and virus-host protein interactions. Human herpesviruses encode distinctively large proteomes, which translate into large numbers of possible intra-viral PPIs. Several studies have attempted to address the description of these networks using a variety of, mostly, experimental approaches [60-63].

The aim of this study was to undertake a systematic and comprehensive analysis of PPI data in HSV1, the prototypical species of human

457 herpesviruses, to characterize the binary and higher order relationships
458 among its encoded proteins. We addressed this using computational
459 approaches to integrate and complement experimental binary PPI data, and
460 analyse its modular structure.
461
462 A first step was to design a new computational framework for reconstructing
463 the PPI network. The three pillars of this framework are: a) data integration; b)
464 standardization; and c) reproducibility. Our method integrates experimentally
465 supported and computationally predicted binary PPI data in a non-redundant
466 manner using standardized molecular interaction data formats. Sequence-
467 based orthology relationships are the basis of our PPI prediction approach.
468 The search for homologues is conducted using the HHblits algorithm, the
469 sensitivity and selectivity of which currently outperforms other available
470 methods [17]. Next, our method applies more conservative criteria than
471 previously [8] to select only the most confident candidate homologues among
472 the results returned by HHblits. Additionally, all interactions are assessed
473 under a common scoring scheme inspired by the standardized MIscore
474 function [18], and which incorporates a new term to penalise the scores of
475 computational predictions in a non-linear way. This term applies to
476 interactions predicted based on protein conservation, and includes information
477 on the prediction method (here, sequence-based interologues mapping), and
478 the number of species from which the interaction is predicted. This aims to
479 reflect that a broader cross-species conservation of protein interactions is
480 indicative of a greater functional relevance of the conservation of the
481 interaction. As a result of our updated, more conservative, protocol, we

observe that although the initial PPI dataset is larger than that used in our previous study (due to additional input databases) [8], the final PPI network is smaller. The reconstructed network is freely available to explore through the newly updated HVint database [8] interface: http://topf-group.ismb.lon.ac.uk/hvint/.

We tested our predictive strategy for the top 5% predicted interactions and we were able to find independent experimental support (i.e. not used to build the network) for 4 out of these 11 interactions. These included PPIs between pUL40 and pUL32, pUL40 and pUL52, pUL5 and pUL52, and the homodimer formed by pUL25 [8,21,22]. This highlights the power of our PPI prediction protocol in proposing new interactions for testing.

The community structure analysis performed on the virion subnetwork of HSV1 indicated the presence of four large communities, consistent with distinct stages of the virion formation process. This analysis suggests new functional relationships among virion components. The delineated communities illustrate events related to capsid formation, early tegumentation at peri- and juxtanuclear regions, and late tegumentation and virion release steps taking place in TGN-derived vesicles and the plasma membrane, respectively.

Community A drew our attention to protein pUS10, so far, a poorly characterized minor component of the virion specific to the α-*Herpesvirinae* subfamily. Our primary sequence analysis of pUS10 predicts the sequence to

be structurally divided into a disordered N-terminus, and an α-helical C-terminus likely to embed a single-pass transmembrane segment. Additionally, our analysis identified a seven-repeat length CLR, centrally in the protein sequence. CLR sequences are characterized by the GXY pattern, where X and Y are any amino acid, and adopt left-handed helical conformations that tend to tightly pack into trimeric right-handed helices [64,65]. CLRs have been identified in a range of organisms, from multicellular eukaryotes to unicellular bacteria and viruses, and they participate in a wide range of processes including e.g. adhesion, morphogenesis, and regulation of signaling cascades among others [64-67]. These features became especially interesting as they are highly reminiscent to those observed in another CLR-containing protein, gp12, from the bacteriophage SPP1 [68,69]. Importantly, SPP1 is one of the species of tailed dsDNA bacteriophages that have been shown to share a common evolutionary ancestry with herpesviruses [70]. Protein gp12 from SPP1 is a capsid-auxiliary protein that contains 8 CLRs centrally located in its sequence, and it is predicted to encode an α-helical C-terminus (Fig 7). The functional oligomeric state of gp12 is a trimer, assembled through the CLR region [69]. This was supported by circular dichroism analyses and confirmed by the cryo-electron microscopy reconstruction of the protein, which also indicated protein flexibility at one end [68,69]. Given the structural similarities with pUS10, it seems reasonable to assume that the functional form of pUS10 is also likely to be a trimer. Gp12 forms spikes that sit at the surface of the capsid, reversibly binding in a temperature-dependent manner, to the centre of the 60 capsid hexons. Binding of gp12 is thought to strengthen the adhesion properties of the virus [69]. Notably, in agreement with this, our

clustering analysis, also predicts close functional association of pUS10 capsid proteins in HSV1.

Fig 7. Structural and functional features of gp12 and pUS10. (A) Schematic comparison of the structural features of gp12 from bacteriophage SPP1 and pUS10 of HSV1. Protein sequences are shown in grey, CLRs as red boxes with the involved residues annotated, and predicted α-helical regions are highlighted as yellow boxes. (B, C) Comparison of the known biological function of gp12 (reversible binding to capsid hexons linked to phage surface adhesion, (B)) and that hypothesised for pUS10 (C). The interaction of pUS10 with HSV1 pUL37, predicted and experimentally supported in this study, would take place in sites of secondary envelopment, such as near Golgi- or trans-Golgi-network (TGN) derived membranes. Accordingly, pUS10 might facilitate the dynamic interactions with membranes in secondary envelopment of capsids through dynamic interactions with the latter, in a similar fashion to gp12's interaction to adhesion surfaces, as well as potentially with membranes by means of its predicted transmembrane segment. Capsids (light blue) in (B) and (C) are drawn approximately to scale.

The identified evolutionary links between capsids of herpesviruses and tailed dsDNA bacteriophages have so far included the conserved HK97-core fold adopted by their major capsid proteins (MCPs), and their general assembly pathways (transitioning from an immature procapsid to a mature, stable, capsid) [68,71,72]. However, no evolutionary ancestry could be traced, based on either sequence or structural features, amongst capsid-associated or capsid auxiliary proteins of the two lineages. Here, we suggest that pUS10

could represent the first of such evolutionary vestiges. Clearly, SPP1 virions differ from those of herpesviruses in that they lack envelope, and, therefore, even in the presence of a common ancestor with gp12, one would expect pUS10 to have partially evolved new functionalities. In this context, the prediction of a transmembrane segment in the C-terminus of pUS10, which is not present in gp12, is interesting. Importantly, our network predicts an interaction between pUS10 and the inner tegument protein pUL37, which was further supported by our immunoaffinity experiments. The results of these experiments demonstrated a higher enrichment of pUS10 co-isolating with pUL37 later in infection (20 hpi). At these time points, pUL37 localises in secondary envelopment sites [37,38]. The functional role of an interaction between pUS10 and pUL37 could be that of facilitating the recruitment of the former to sites of secondary envelopment, to cooperatively function in subsequent capsid tegumentation events (Fig 7). An association of pUS10 to capsids via pUL37 would also explain why pUS10 has not yet been observed in the recent high-resolution structures of herpesviral capsids, [73-75]. On the basis of the structural similarities with gp12, it is reasonable to think that pUS10 could also exhibit reversible capsid/inner tegument-binding properties, similar to gp12, and establish dynamic interactions with capsid or capsid-associated proteins whilst, at the same time, interact with Golgi- or trans-Golgi-derived membranes, at the prospective capsids proximal pole of the forming virion, through its predicted transmembrane segment, and promote the final step in capsid trafficking during virion morphogenesis.

Together with pUS10, 9 out of the 14 proteins that our network data predict interact with pUL37, were co-purified with the latter in our immunoaffinity purification experiments (Fig 6D). Although these data cannot conclusively answer which binary interactions are direct ones among the co-isolated proteins, they add further support to the likely functional association of these proteins, and therefore add to the confidence of the PPI involved in the pUL37-centered subnetwork. Other interesting features of pUS10 will surely help defining its functionalities. For instance, the polyproline sequence located between two of the predicted C-terminus α-helices is also likely to be involved in PPI [76].

Although further investigations on the roles of pUS10 during viral infection are required, our analysis has revealed insightful new data on the structural features and potential evolutionary history of this intriguing and, so far understudied protein. Combined, our analyses have allowed to pose functional hypotheses that will further our understanding of the intricate tegumentation process of herpesvirus virions.

A second community sheds light to early tegumentation stages, specifically to the recruitment of a group of transcription and translation modulators, playing key roles at early stages of the lytic life cycle, into the virion particle. pUL48, also called VP16 or α-TIF, is a transcription activator that migrates with capsids to the host nucleus upon infection, and activates transcription of lytic genes [77]. The roles of pUL47 and pUL49 are more diverse, however both of them, together with pUL48 participate in regulating the activity of pUL41, also

known as the virus host shut-off (VHS) protein, which inhibits protein synthesis [78]. Additionally, binary interactions among the four proteins seem to be required for their incorporation to the virion [39,40,79,80]. Yet, the stage at which this happens and the events that lead the process are unclear. Our results suggest that the recruitment of this module into the particle could be guided by the coordinated action of proteins pUL16, pUL11, pUL21, pUL14 and pUL56, potentially in peri- or juxtanuclear regions.

Two other communities are enriched in late stages of the tegumentation process and virion release. We observed a large number of proteins annotated with trafficking of TGN-derived vesicles and membrane-regulatory events. Particularly, our analysis suggests a strong clustering tendency between gK/UL20, pUL45, pUS2, and the poorly characterized *α-Herpesvirinae*-specific protein pUL55 [53]. Primary sequence analysis indicated pUL55 is a globular α/β protein with no predicted transmembrane regions. However, all four proteins gK, pUL20, pUL45, and pUS2, are membrane or membrane-associated proteins during intracellular stages of the virus life cycle, and gK, pUL20, and pUL45 are also envelope-associated proteins [54,81,82]. This finding supports the idea that pUL55 is a component of the outer most layer of the tegument, and its recruitment might be dependent on gK/UL20, pUL45, pUS2.

The outcomes of our protocol for predicting both binary and higher-order protein associations, bring novel insights and direction to future experimental testing. Further, we are currently working on the implementation of the

introduced network assembly and analysis frameworks on other species of human herpesviruses. Widening this analysis to other members of this important group of human pathogens will also underscore both conserved and species-specific features of their interactome organization, and help to explain the observed phenotypes and evolution of their pathogenic strategies.

Materials and Methods

Data collection, curation, and integration

Binary PPI data from molecular interaction repositories were downloaded in the standardised MITAB 2.5 format [20] (Fig 1). Binary PPIs from PDB and EMDB entries were manually extracted. Only entries with assigned PubMed identifier were considered. In the case of EMDB entries, only those with resolutions ≤5Å at the time of this work were taken into account. When fitted atomic models were available, binary interactions between two proteins were extracted if the interface between them was larger than 500Å^2 [83-85]. When fitted atomic models were not available, binary interactions were assigned as described in the primary citation. The resulting set of interactions were next manually annotated following the MITAB 2.5 format.

The data were collected for the species in S1 and S2 Tables. Only binary PPIs for which both interacting proteins were annotated with a taxonomic identifier corresponding to these species (all strains considered) were selected. Taxonomic identifiers were obtained from the National Center for

Biotechnology Information (NCBI) [86] Taxonomy database (https://www.ncbi.nlm.nih.gov/taxonomy).

Protein identifiers were mapped, where possible, to UniProtKB [7] accession numbers. PPIs for which such mapping was not possible for one or both of the interacting proteins, were not further considered in our pipeline. The seven input PPI data sets were merged into a single non-redundant collection.

Computational prediction of PPIs

PPI predictions were obtained using an interologues mapping approach [16]. This method (Fig 1D) predicts an interaction in species X if two proteins, known to interact in species Y, are conserved in both species X and Y. Here, protein conservation was assessed using sequence-based orthology predictions computed with the iterative Hidden Markov Model (HMM) profile comparison algorithm implemented by HHblits [17]. For each query sequence, the best match found in HSV1 satisfying all of the following conditions was considered a reliable putative homologue: ≥20% sequence identity, ≥30% sequence similarity, ≥50% query HMM profile coverage, and ≥95% probability of being a true positive. The resulting set of candidate homology relationships were used to infer interologues in each target interactomes.

Integration of validated and predicted PPIs

Experimentally validated and computationally predicted interactions were merged into a single interactome data set. Strain redundancy was removed by mapping all protein sequences to reference strain accession numbers (HSV1 strain 17) using UniRef90 clusters.

PPI scoring function

The scoring scheme integrated in our framework is inspired by the standardised MIscore function [18]. Under this scheme, PPIs that had experimental support in the target species (with or without additional support from computational predictions) were scored using the MIscore function. This was done through the MImerge service [18]. PPIs that did not have experimental support were scored with a new scoring function, defined as in Equation 1.

$$Score = MIscore \times (Penalty\ function) \quad \text{(Equation 1)}$$

The new scoring function consists of first scoring an interaction using the MIscore [18], and next applying a penalty function to the returned score. This scaling factor (Equation 2) takes as reference the structure of the terms used in the MIscore function, but it redefines the meaning of their parameters to incorporate information on the number of species and prediction method used.

$$Scaling\ factor = \log_{(2a+1)}(a+1) \quad \text{(Equation 2)}$$

where

$$a = scv \times n$$

where scv is the score associated to the PPI prediction method (in this case interology mapping), and n is the number of different species from which the interaction was predicted.

The value of the penalty function increases asymptotically from ~0.5 to 1 with the number of species from which an interaction is predicted. Because we considered 10 orthologues to each target species, the values of the scaling factor in this study fall in the [0.5, 0.6] range. After applying the penalty function the values are normalised within [0, 1].

Consensus clustering framework

Prior to clustering, the signal-to-noise ratio in the network data was increased as follows (Fig 3A). First, the subnetwork formed by proteins found in the mature virion particle of HSV1 (i.e. HSV1 virion subnetwork) was extracted from the full network. Next, self-interactions were removed as they do not add topological information about the network. Interactions between capsid and envelope proteins were also eliminated as they are precluded by the presence of the tegument in between the two former layers. Finally, the LCC was obtained and used as input to the consensus clustering framework.

A total of 14 different clustering algorithms, representing a broad spectrum of clustering techniques, were then applied to the LCC (Fig 3B), i.e. the K-means algorithm [87], agglomerative hierarchical clustering [88], K-means/hierarchical clustering [89], Fuzzy C-means [90], Model-based clustering [91,92], Markov Cluster algorithm (MCL) algorithm [93], Density-based clustering [94], Edge betweenness [95,96], Louvain method [97], and Leading eigenvector [98], Fast greedy [28], Walktrap [99], and InfoMap [100] algorithms. Where needed, partition selection and parameter optimisation were conducted based on modularity [95] maximisation. The clustering results of those partitions yielding positive modularity were integrated into a consensus agreement matrix (CAM) (Fig 3C and S4 Table). The condition of positive modularity discarded partitions were each and every node was singled out as a cluster by itself, the entire network was treated as a super-cluster, or the graph under such partition had less local structure than random models. The columns and rows of the consensus agreement matrix represented the network proteins; the matrix values indicated the fraction of accepted base partitions classifying two proteins in the same cluster.

A final consensus partition was derived from the CAM by applying a filtering threshold (Fig 3D and S5 Table). For thresholds between 0 and 1, by increases of 0.5, we iteratively calculated a consensus partition in the following way. CAM cells with values equal to or below the threshold were ignored; the remaining cells defined a partition for which the modularity was calculated. At the end of this iterative process, the partition associated to the highest modularity was chosen as consensus.

752

753

754 **Sequence analysis**

755

756 The canonical sequence of pUS10 and pUL55 in the HSV1 reference
757 proteome were obtained from UniProtKB [7]. For each sequence, the
758 following analysis was conducted. ScanProsite [101] was used to scan the
759 sequences for sequence motifs. Potential sequence homologues within the
760 entire UniProtKB database [7], were searched for using HHblits [17]. Next, a
761 consensus prediction of secondary structure elements was inferred from the
762 results of four different secondary structure prediction tools, i.e. SPIDER^2
763 [102], PSIPRED [103], JPred4 [104], and PSSpred [105]. Probabilities
764 associated to the returned predictions were not integrated in the consensus
765 analysis. Similarly, consensus predictions for transmembrane segments were
766 derived from five different algorithms, i.e. Dense Alignment Surface (DAS)
767 [106], Phobius [107], PHDhtml [108], TMpred [109], and MEMSAT-SVM [110].
768 From TMpred predictions only significant regions (defined as regions with
769 score above 500) and core residues were taken into consideration. From
770 Phobius, only residues with probability of belonging to a transmembrane
771 region above 0.1 were considered. Finally, a consensus prediction for
772 disordered regions was built from the results of two algorithms, i.e.
773 DISOPRED [111] and MetaDisorder [112].

774

775 **Infection of human fibroblasts with HSV1 and live cell imaging**

We used the HSV1(17^+)Lox-UL37GFP strain, a generous gift from B. Sodeik and previously characterized in [37], here denoted HSV1-UL37EGFP, and as a control, the HSV1(17^+)Lox-P_{MCMV}GFP strain, denoted HSV1-EGFP, which expresses EGFP alone inserted between the pUL55 and pUL56 ORFs, under the control of the murine cytomegalovirus promoter [113]. Viruses were propagated, isolated, and titered in Vero cells (ATCC CCL81) grown in DMEM containing 10% FBS and 1% penicillin/streptomycin (P/S), as previously described [8]. Primary human foreskin fibroblast cells were infected with HSV-1 strains at 10 plaque forming units/cell using a cold-synchronized protocol [114]. The progression of infection was visualized by live-cell imaging on a Nikon Ti-Eclipse epifluorescence inverted microscope from 2 hpi to 24 hpi. Images were viewed and analysed by ImageJ [115].

Immunoaffinity Purification Quantitative Mass Spectrometry

For IP-MS experiments, cells were infected as above with HSV1-UL37GFP or control HSV1-GFP, in duplicate. Infected cells were collected at 8 and 20 hpi (HSV1-UL37GFP) or 20 hpi (HSV1-GFP) in ice-cold PBS and pelleted by centrifugation (~1 x 10^7 per time point per replicate). Cell pellets were washed in ice-cold PBS and lysed hypotonically. Cytosolic lysates were adjusted to 20 mM HEPES-KOH, pH 7.4, containing 0.11 M potassium acetate, 2 mM $MgCl_2$, 0.1% Tween 20, 1 µM $ZnCl_2$, 1 µM $CaCl_2$, 250 mM NaCl, and 0.5% NP-40, mixed by Polytron homogenization, and centrifuged at 8,000 x g for 10 min at 4°C. The supernatant was recovered and subjected to immunoaffinity

purification using magnetic beads conjugated with in-house generated rabbit anti-GFP antibodies, as previously described [114,116].

Immunoisolated proteins were processed by a Filter-Aided Sample Preparation method using Amicon ultrafiltration devices (Millipore, 30 kDa MWCO) as described [117], except 0.1 M Tris-HCl, pH 7.9 was replaced with 0.1 M triethylammonium bicarbonate (TEAB). Following overnight trypsin digestion and clean-up, peptides (4 µl) were analyzed by nanoliquid chromatography-tandem mass spectrometry on a Dionex Ultimate 3000 RSLC coupled directly to a LTQ Orbitrap Velos ETD configured with a Nanospray ion source (Thermofisher Scientific).

The Proteome Discoverer software (ver. 2.2) was used for post-acquisition mass recalibration of precursor and fragment ions masses, MS/MS spectrum extraction, peptide spectrum matching and validation, calculation of TMT reporter ion intensities, and assembly of quantified into protein groups. Protein groups and TMT protein abundances for herpesvirus proteins with a minimum of 2 unique quantified peptides were exported to Excel. IP protein enrichment ratios for each time point and replicate were calculated as the TMT abundance ratio of pUL37GFP / GFP. Proteins with IP enrichment ratios of ≥ 2-fold in at least one time point in both replicates were considered specific associations. The TMT abundance ratio for proteins in the 20 vs 8 hpi pUL37GFP IPs were calculated after normalization by the pUL37 TMT abundance. Further details on data collection and analysis can be found in S1 Text.

Acknowledgments

We are grateful for funding from Human Frontiers Science Program (RGY0079/2009-C) to K.G., M.T., and I.M.C., from the NIH (GM114141) and a Mallinckrodt Scholar Award to I.M.C., from the Wellcome Trust (209250/Z/17/Z) to M.T. and K.G., and (107806/Z/15/Z) and a WT Core Award (203141/Z/16/Z) to K.G., and from the MRC (MR/M019292/1) to M.T. and K.G.

References

1. Arvin A, Campadelli-Fiume G, Mocarski E, Moore PS, Roizman B, Whitley R, et al. Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis. Cambridge: Cambridge University Press; 2007.

2. Eimer WA, Vijaya Kumar DK, Navalpur Shanmugam NK, Rodriguez AS, Mitchell T, Washicosky KJ, et al. Alzheimer's Disease-Associated β-Amyloid Is Rapidly Seeded by Herpesviridae to Protect against Brain Infection. Neuron. 2018;99: 56–63.e3. doi:10.1016/j.neuron.2018.06.030

3. Readhead B, Haure-Mirande J-V, Funk CC, Richards MA, Shannon P, Haroutunian V, et al. Multiscale Analysis of Independent Alzheimer's Cohorts Finds Disruption of Molecular, Genetic, and Clinical Networks by Human Herpesvirus. Neuron. Cell Press; 2018;99: 64–82.e7. doi:10.1016/j.neuron.2018.05.023

4. Grünewald K, Desai P, Winkler DC, Heymann JB, Belnap DM, Baumeister W, et al. Three-Dimensional Structure of Herpes Simplex Virus from Cryo-Electron Tomography. Science. American Association for the Advancement of Science; 2003;302: 1396–1398. doi:10.1126/science.1090284

5. Spear PG, Longnecker R. Herpesvirus entry: an update. J Virol. American Society for Microbiology (ASM); 2003;77: 10179–10185. doi:10.1128/JVI.77.19.10179-10185.2003

6. Arvin A, Campadelli-Fiume G, Mocarski E, Moore PS, Roizman B, Whitley R, et al. Comparative virion structures of human herpesviruses. Cambridge: Cambridge University Press; 2007.

7. Consortium TU. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 2017;45: D158–D169. doi:10.1093/nar/gkw1099

8. Ashford P, Hernandez A, Greco TM, Buch A, Sodeik B, Cristea IM, et al. HVint: A Strategy for Identifying Novel Protein-Protein Interactions in Herpes Simplex Virus Type 1. Mol Cell Proteomics. American Society for Biochemistry and Molecular Biology; 2016;15: 2939–2953. doi:10.1074/mcp.M116.058552

9. Stark C, Breitkreutz B-J, Reguly T, Boucher L, Breitkreutz A, Tyers M. BioGRID: a general repository for interaction datasets. Nucleic Acids Res. 2006;34: D535–9. doi:10.1093/nar/gkj109

10. Xenarios I, Rice DW, Salwinski L, Baron MK, Marcotte EM, Eisenberg D. DIP: the database of interacting proteins. Nucleic

Acids Res. 2000;28: 289–291.

11. Orchard S, Ammari M, Aranda B, Breuza L, Briganti L, Broackes-Carter F, et al. The MIntAct project--IntAct as a common curation platform for 11 molecular interaction databases. Nucleic Acids Res. 2014;42: D358–63. doi:10.1093/nar/gkt1115

12. Calderone A, Castagnoli L, Cesareni G. mentha: a resource for browsing integrated protein-interaction networks. Nat Methods. Nature Publishing Group; 2013;10: 690–691. doi:10.1038/nmeth.2561

13. Guirimand T, Delmotte S, Navratil V. VirHostNet 2.0: surfing on the web of virus/host molecular interactions data. Nucleic Acids Res. 2015;43: D583–7. doi:10.1093/nar/gku1121

14. Berman H, Henrick K, Nakamura H. Announcing the worldwide Protein Data Bank. Nat Struct Mol Biol. Nature Publishing Group; 2003;10: 980–980. doi:10.1038/nsb1203-980

15. Lawson CL, Baker ML, Best C, Bi C, Dougherty M, Feng P, et al. EMDataBank.org: unified data resource for CryoEM. Nucleic Acids Res. 2011;39: D456–64. doi:10.1093/nar/gkq880

16. Yu H. Annotation Transfer Between Genomes: Protein-Protein Interologs and Protein-DNA Regulogs. Genome Research. Cold Spring Harbor Lab; 2004;14: 1107–1118. doi:10.1101/gr.1774904

17. Remmert M, Biegert A, Hauser A, Söding J. HHblits: lightning-fast

iterative protein sequence searching by HMM-HMM alignment. Nat Methods. Nature Publishing Group; 2011;9: 173–175. doi:10.1038/nmeth.1818

18. Villaveces JM, Jiménez RC, Porras P, Del-Toro N, Duesbury M, Dumousseau M, et al. Merging and scoring molecular interactions utilising existing community standards: tools, use-cases and a case study. Database (Oxford). 2015;2015: bau131–bau131. doi:10.1093/database/bau131

19. Orchard S, Hermjakob H, Apweiler R. The proteomics standards initiative. Proteomics. Wiley-Blackwell; 2003;3: 1374–1376. doi:10.1002/pmic.200300496

20. Hermjakob H, Montecchi-Palazzi L, Bader G, Wojcik J, Salwinski L, Ceol A, et al. The HUPO PSI's molecular interaction format--a community standard for the representation of protein interaction data. Nat Biotechnol. Nature Publishing Group; 2004;22: 177–183. doi:10.1038/nbt926

21. Biswas N, Weller SK. The UL5 and UL52 subunits of the herpes simplex virus type 1 helicase-primase subcomplex exhibit a complex interdependence for DNA binding. J Biol Chem. American Society for Biochemistry and Molecular Biology; 2001;276: 17610–17619. doi:10.1074/jbc.M010107200

22. Dai X, Zhou ZH. Structure of the herpes simplex virus 1 capsid with associated tegument protein complexes. Science. American

Association for the Advancement of Science; 2018;360: eaao7298. doi:10.1126/science.aao7298

23. Bechtel JT, Winant RC, Ganem D. Host and Viral Proteins in the Virion of Kaposi's Sarcoma-Associated Herpesvirus. J Virol. American Society for Microbiology Journals; 2005;79: 4952–4964. doi:10.1128/JVI.79.8.4952-4964.2005

24. Loret S, Guay G, Lippé R. Comprehensive characterization of extracellular herpes simplex virus type 1 virions. J Virol. American Society for Microbiology Journals; 2008;82: 8605–8618. doi:10.1128/JVI.00904-08

25. Human cytomegalovirus virion proteins. Human Immunology. Elsevier; 2004;65: 395–402. doi:10.1016/j.humimm.2004.02.008

26. Johannsen E, Luftig M, Chase MR, Weicksel S, Cahir-McFarland E, Illanes D, et al. Proteins of purified Epstein-Barr virus. PNAS. National Academy of Sciences; 2004;101: 16286–16291. doi:10.1073/pnas.0407320101

27. Newman MEJ, Girvan M. Finding and evaluating community structure in networks. Phys Rev E Stat Nonlin Soft Matter Phys. American Physical Society; 2004;69: 026113. doi:10.1103/PhysRevE.69.026113

28. Clauset A, Newman MEJ, Moore C. Finding community structure in very large networks. Phys Rev E Stat Nonlin Soft Matter Phys. American Physical Society; 2004;70: 066111.

doi:10.1103/PhysRevE.70.066111

29. research GOCNA, 2016. Expansion of the Gene Ontology knowledgebase and resources. academicoupcom

.

30. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. Nature Publishing Group; 2000;25: 25–29. doi:10.1038/75556

31. Kornfeind EM, Visalli RJ. Human herpesvirus portal proteins: Structure, function, and antiviral prospects. Reviews in Medical Virology. Wiley-Blackwell; 2018;28: e1972. doi:10.1002/rmv.1972

32. Yamada H, Daikoku T, Yamashita Y, Jiang YM, Tsurumi T, Nishiyama Y. The product of the US10 gene of herpes simplex virus type 1 is a capsid/tegument-associated phosphoprotein which copurifies with the nuclear matrix. J Gen Virol. Microbiology Society; 1997;78 (Pt 11): 2923–2931. doi:10.1099/0022-1317-78-11-2923

33. Roger Holden V, Yalamanchili RR, Harty RN, O'Callaghan DJ. Identification and characterization of an equine herpesvirus 1 late gene encoding a potential zinc finger. Virology. Academic Press; 1992;188: 704–713. doi:10.1016/0042-6822(92)90525-T

34. Charalambous BM, Keen JN, McPherson MJ. Collagen-like

960 sequences stabilize homotrimers of a bacterial hydrolase. EMBO J.
961 European Molecular Biology Organization; 1988;7: 2903–2909.

962 35. Desai P, Sexton GL, McCaffery JM, Person S. A null mutation in
963 the gene encoding the herpes simplex virus type 1 UL37
964 polypeptide abrogates virus maturation. J Virol. American Society
965 for Microbiology; 2001;75: 10259–10271.
966 doi:10.1128/JVI.75.21.10259-10271.2001

967 36. Pasdeloup D, Beilstein F, Roberts APE, McElwee M, McNab D,
968 Rixon FJ. Inner tegument protein pUL37 of herpes simplex virus
969 type 1 is involved in directing capsids to the trans-Golgi network for
970 envelopment. Journal of General Virology. Microbiology Society;
971 2010;91: 2145–2151. doi:10.1099/vir.0.022053-0

972 37. Sandbaumhüter M, Döhner K, Schipke J, Binz A, Pohlmann A,
973 Sodeik B, et al. Cytosolic herpes simplex virus capsids not only
974 require binding inner tegument protein pUL36 but also pUL37 for
975 active transport prior to secondary envelopment. Cell Microbiol.
976 Wiley/Blackwell (10.1111); 2013;15: 248–269.
977 doi:10.1111/cmi.12075

978 38. Desai P, Sexton GL, Huang E, Person S. Localization of herpes
979 simplex virus type 1 UL37 in the Golgi complex requires UL36 but
980 not capsid structures. J Virol. American Society for Microbiology
981 Journals; 2008;82: 11354–11361. doi:10.1128/JVI.00956-08

982 39. Zhang Y, Sirko DA, McKnight JL. Role of herpes simplex virus type

1 UL46 and UL47 in alpha TIF-mediated transcriptional induction: characterization of three viral deletion mutants. J Virol. American Society for Microbiology; 1991;65: 829–841.

40. Smibert CA, Popova B, Xiao P, Capone JP, Smiley JR. Herpes simplex virus VP16 forms a complex with the virion host shutoff protein vhs. J Virol. American Society for Microbiology (ASM); 1994;68: 2339–2346.

41. Elliott G, Hafezi W, Whiteley A, Bernard E. Deletion of the herpes simplex virus VP22-encoding gene (UL49) alters the expression, localization, and virion incorporation of ICP0. J Virol. 2005;79: 9735–9745. doi:10.1128/JVI.79.15.9735-9745.2005

42. Klupp BG, Böttcher S, Granzow H, Kopp M, Mettenleiter TC. Complex formation between the UL16 and UL21 tegument proteins of pseudorabies virus. J Virol. American Society for Microbiology Journals; 2005;79: 1510–1522. doi:10.1128/JVI.79.3.1510-1522.2005

43. Yeh P-C, Meckes DG, Wills JW. Analysis of the interaction between the UL11 and UL16 tegument proteins of herpes simplex virus. J Virol. American Society for Microbiology Journals; 2008;82: 10693–10700. doi:10.1128/JVI.01230-08

44. Harper AL, Meckes DG, Marsh JA, Ward MD, Yeh P-C, Baird NL, et al. Interaction domains of the UL16 and UL21 tegument proteins of herpes simplex virus. J Virol. American Society for Microbiology;

2010;84: 2963–2971. doi:10.1128/JVI.02015-09

45. Chadha P, Han J, Starkey JL, Wills JW. Regulated interaction of tegument proteins UL16 and UL11 from herpes simplex virus. J Virol. American Society for Microbiology Journals; 2012;86: 11886–11898. doi:10.1128/JVI.01879-12

46. Tunbäck P, Liljeqvist JA, Löwhagen GB, Bergström T. Glycoprotein G of herpes simplex virus type 1: identification of type-specific epitopes by human antibodies. J Gen Virol. Microbiology Society; 2000;81: 1033–1040. doi:10.1099/0022-1317-81-4-1033

47. Klupp BG, Altenschmidt J, Granzow H, Fuchs W, Mettenleiter TC. Identification and characterization of the pseudorabies virus UL43 protein. Virology. 2005;334: 224–233. doi:10.1016/j.virol.2005.01.032

48. Aubert M, Krantz EM, Jerome KR. Herpes simplex virus genes Us3, Us5, and Us12 differentially regulate cytotoxic T lymphocyte-induced cytotoxicity. Viral Immunol. Mary Ann Liebert, Inc. 2 Madison Avenue Larchmont, NY 10538 USA; 2006;19: 391–408. doi:10.1089/vim.2006.19.391

49. Aubert M, Chen Z, Lang R, Dang CH, Fowler C, Sloan DD, et al. The antiapoptotic herpes simplex virus glycoprotein J localizes to multiple cellular organelles and induces reactive oxygen species formation. J Virol. American Society for Microbiology Journals; 2008;82: 617–629. doi:10.1128/JVI.01341-07

50. Dollery SJ, Lane KD, Delboy MG, Roller DG, Nicola AV. Role of the UL45 protein in herpes simplex virus entry via low pH-dependent endocytosis and its relationship to the conformation and function of glycoprotein B. Virus Res. 2010;149: 115–118. doi:10.1016/j.virusres.2010.01.004

51. Kasmi El I, Lippé R. Herpes Simplex Virus 1 gN Partners with gM To Modulate the Viral Fusion Machinery. Longnecker RM, editor. J Virol. American Society for Microbiology; 2015;89: 2313–2323. doi:10.1128/JVI.03041-14

52. Lau S-YK, Crump CM. HSV-1 gM and the gK/pUL20 complex are important for the localization of gD and gH/L to viral assembly sites. Viruses. Multidisciplinary Digital Publishing Institute; 2015;7: 915–938. doi:10.3390/v7030915

53. Yamada H, Jiang YM, Oshima S, Daikoku T, Yamashita Y, Tsurumi T, et al. Characterization of the UL55 gene product of herpes simplex virus type 2. Journal of General Virology. Microbiology Society; 1998;79: 1989–1995. doi:10.1099/0022-1317-79-8-1989

54. Kang M-H, Roy BB, Finnen RL, Le Sage V, Johnston SM, Zhang H, et al. The Us2 gene product of herpes simplex virus 2 is a membrane-associated ubiquitin-interacting protein. J Virol. 2013;87: 9590–9603. doi:10.1128/JVI.00994-13

55. Roller RJ, Haugo AC, Yang K, Baines JD. The herpes simplex virus 1 UL51 gene product has cell type-specific functions in cell-to-cell

spread. J Virol. 2014;88: 4058–4068. doi:10.1128/JVI.03707-13

56. Albecka A, Owen DJ, Ivanova L, Brun J, Liman R, Davies L, et al. Dual Function of the pUL7-pUL51 Tegument Protein Complex in Herpes Simplex Virus 1 Infection. Sandri-Goldin RM, editor. J Virol. American Society for Microbiology Journals; 2017;91: 448. doi:10.1128/JVI.02196-16

57. Lu X, Huang C, Zhang Y, Lin Y, Wang X, Li Q, et al. The Us2 Gene Product of Herpes Simplex Virus 2 modulates NF-κB activation by targeting TAK1. Sci Rep. Nature Publishing Group; 2017;7: 8396. doi:10.1038/s41598-017-08856-4

58. Dingwell KS, Johnson DC. The herpes simplex virus gE-gI complex facilitates cell-to-cell spread and binds to components of cell junctions. J Virol. American Society for Microbiology; 1998;72: 8933–8942.

59. DuRaine G, Wisner TW, Howard P, Williams M, Johnson DC. Herpes Simplex Virus gE/gI and US9 Promote both Envelopment and Sorting of Virus Particles in the Cytoplasm of Neurons, Two Processes That Precede Anterograde Transport in Axons. Longnecker RM, editor. J Virol. American Society for Microbiology Journals; 2017;91: 153. doi:10.1128/JVI.00050-17

60. Fossum E, Friedel CC, Rajagopala SV, Titz B, Baiker A, Schmidt T, et al. Evolutionarily conserved herpesviral protein interaction networks. Sun R, editor. PLoS Pathog. Public Library of Science;

2009;5: e1000570. doi:10.1371/journal.ppat.1000570

61. Vittone V, Diefenbach E, Triffett D, Douglas MW, Cunningham AL, Diefenbach RJ. Determination of interactions between tegument proteins of herpes simplex virus type 1. J Virol. American Society for Microbiology Journals; 2005;79: 9566–9571. doi:10.1128/JVI.79.15.9566-9571.2005

62. Lee JH, Vittone V, Diefenbach E, Cunningham AL, Diefenbach RJ. Identification of structural protein–protein interactions of herpes simplex virus type 1. Virology. Academic Press; 2008;378: 347–354. doi:10.1016/j.virol.2008.05.035

63. Calderwood MA, Venkatesan K, Xing L, Chase MR, Vazquez A, Holthaus AM, et al. Epstein-Barr virus and virus human protein interaction maps. PNAS. National Academy of Sciences; 2007;104: 7606–7611. doi:10.1073/pnas.0702332104

64. Brodsky B, Ramshaw JA. The collagen triple-helix structure. Matrix Biol. 1997;15: 545–554.

65. Yu Z, An B, Ramshaw JAM, Brodsky B. Bacterial collagen-like proteins that form triple-helical structures. Journal of Structural Biology. 2014;186: 451–461. doi:10.1016/j.jsb.2014.01.003

66. Rasmussen M, Jacobsson M, Björck L. Genome-based Identification and Analysis of Collagen-related Structural Motifs in Bacterial and Viral Proteins. J Biol Chem. American Society for Biochemistry and Molecular Biology; 2003;278: 32313–32316.

doi:10.1074/jbc.M304709200

67. Mienaltowski MJ, Birk DE. Structure, physiology, and biochemistry of collagens. Adv Exp Med Biol. Dordrecht: Springer Netherlands; 2014;802: 5–29. doi:10.1007/978-94-007-7893-1_2

68. White HE, Sherman MB, Brasilès S, Jacquet E, Seavers P, Tavares P, et al. Capsid structure and its stability at the late stages of bacteriophage SPP1 assembly. J Virol. American Society for Microbiology Journals; 2012;86: 6768–6777. doi:10.1128/JVI.00412-12

69. Zairi M, Stiege AC, Nhiri N, Jacquet E, Tavares P. The collagen-like protein gp12 is a temperature-dependent reversible binder of SPP1 viral capsids. J Biol Chem. American Society for Biochemistry and Molecular Biology; 2014;289: 27169–27181. doi:10.1074/jbc.M114.590877

70. Baker ML, Jiang W, Rixon FJ, Chiu W. Common ancestry of herpesviruses and tailed DNA bacteriophages. J Virol. American Society for Microbiology Journals; 2005;79: 14967–14970. doi:10.1128/JVI.79.23.14967-14970.2005

71. Gertsman I, Gan L, Guttman M, Lee K, Speir JA, Duda RL, et al. An unexpected twist in viral capsid maturation. Nature. 2009;458: 646–650. doi:10.1038/nature07686

72. Parent KN, Khayat R, Tu LH, Suhanovsky MM, Cortines JR, Teschke CM, et al. P22 coat protein structures reveal a novel

mechanism for capsid maturation: stability without auxiliary proteins or chemical crosslinks. Structure. 2010;18: 390–401. doi:10.1016/j.str.2009.12.014

73. Dai X, Zhou ZH. Structure of the herpes simplex virus 1 capsid with associated tegument protein complexes. Science. American Association for the Advancement of Science; 2018;360: eaao7298. doi:10.1126/science.aao7298

74. Yu X, Jih J, Jiang J, Zhou ZH. Atomic structure of the human cytomegalovirus capsid with its securing tegument layer of pp150. Science. American Association for the Advancement of Science; 2017;356: eaam6892. doi:10.1126/science.aam6892

75. Dai X, Gong D, Lim H, Jih J, Wu T-T, Sun R, et al. Structure and mutagenesis reveal essential capsid protein interactions for KSHV replication. Nature. Nature Publishing Group; 2018;553: 521–525. doi:10.1038/nature25438

76. Berisio R, Vitagliano L. Polyproline and triple helix motifs in host-pathogen recognition. Curr Protein Pept Sci. Bentham Science Publishers; 2012;13: 855–865. doi:10.2174/138920312804871157

77. Wysocka J, Herr W. The herpes simplex virus VP16-induced complex: the makings of a regulatory switch. Trends Biochem Sci. 2003;28: 294–304. doi:10.1016/S0968-0004(03)00088-4

78. Smiley JR. Herpes simplex virus virion host shutoff protein: immune evasion mediated by a viral RNase? J Virol. American Society for

Microbiology (ASM); 2004;78: 1063–1068. doi:10.1128/JVI.78.3.1063-1068.2004

79. Elliott G, Mouzakitis G, O'Hare P. VP16 interacts via its activation domain with VP22, a tegument protein of herpes simplex virus, and is relocated to a novel macromolecular assembly in coexpressing cells. J Virol. American Society for Microbiology (ASM); 1995;69: 7932–7941.

80. Taddeo B, Sciortino MT, Zhang W, Roizman B. Interaction of herpes simplex virus RNase with VP16 and VP22 is required for the accumulation of the protein but not for accumulation of mRNA. PNAS. National Academy of Sciences; 2007;104: 12163–12168. doi:10.1073/pnas.0705245104

81. Dollery SJ, Lane KD, Delboy MG, Roller DG, Nicola AV. Role of the UL45 protein in herpes simplex virus entry via low pH-dependent endocytosis and its relationship to the conformation and function of glycoprotein B. Virus Res. Elsevier; 2010;149: 115–118. doi:10.1016/j.virusres.2010.01.004

82. Lau S-YK, Crump CM. HSV-1 gM and the gK/pUL20 complex are important for the localization of gD and gH/L to viral assembly sites. Viruses. Multidisciplinary Digital Publishing Institute; 2015;7: 915–938. doi:10.3390/v7030915

83. Jones S, Thornton JM. Principles of protein-protein interactions. PNAS. National Academy of Sciences; 1996;93: 13–20.

84. Bogan AA, Thorn KS. Anatomy of hot spots in protein interfaces. J Mol Biol. 1998;280: 1–9. doi:10.1006/jmbi.1998.1843

85. Day ES, Cote SM, Whitty A. Binding efficiency of protein-protein complexes. Biochemistry. American Chemical Society; 2012;51: 9124–9136. doi:10.1021/bi301039t

86. Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, et al. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2006;34: D173–80. doi:10.1093/nar/gkj158

87. MacQueen J. Some methods for classification and analysis of multivariate observations. Berkeley, Calif.: University of California Press; 1967. pp. 281–297.

88. Franklin J. The elements of statistical learning: data mining, inference and prediction. The Mathematical Intelligencer. Springer-Verlag; 2008;27: 83–85. doi:10.1007/BF02985802

89. Bernard Chen, Harrison R, Yi Pan, Phang C Tai. Novel Hybrid Hierarchical-K-means Clustering Method (H-K-means) for Microarray Analysis. IEEE; pp. 105–108. doi:10.1109/CSBW.2005.98

90. Liu H-C, Wu D-B, Yih J-M, Liu S-W. Fuzzy c-Mean Algorithm Based on Complete Mahalanobis Distances and Separable Criterion. IEEE; 2008. pp. 87–91. doi:10.1109/FSKD.2008.34

91. Banfield JD, Raftery AE. Model-Based Gaussian and Non-Gaussian Clustering. Biometrics. 1993;49: 803. doi:10.2307/2532201

92. Fraley C, Raftery AE. MCLUST: Software for Model-Based Cluster Analysis. Journal of Classification. Springer-Verlag; 2014;16: 297–306. doi:10.1007/s003579900058

93. Enright AJ, Van Dongen S, Ouzounis CA. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. Oxford University Press; 2002;30: 1575–1584.

94. Ester M, Kriegel H-P, Sander JOR, Xu X. A Density-based Algorithm for Discovering Clusters a Density-based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. AAAI Press; 1996. pp. 226–231.

95. Newman MEJ, Girvan M. Finding and evaluating community structure in networks. Phys Rev E Stat Nonlin Soft Matter Phys. American Physical Society; 2004;69: 268. doi:10.1103/PhysRevE.69.026113

96. Girvan M, Newman MEJ. Community structure in social and biological networks. PNAS. National Academy of Sciences; 2002;99: 7821–7826. doi:10.1073/pnas.122653799

97. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J Stat Mech. IOP Publishing; 2008;2008: P10008. doi:10.1088/1742-5468/2008/10/P10008

98. Newman MEJ. Finding community structure in networks using the eigenvectors of matrices. Phys Rev E Stat Nonlin Soft Matter Phys. American Physical Society; 2006;74: 036104. doi:10.1103/PhysRevE.74.036104

99. Pons P, Latapy M. Computing Communities in Large Networks Using Random Walks. Computer and Information Sciences - ISCIS 2005. Berlin, Heidelberg: Springer Berlin Heidelberg; 2005. pp. 284–293. doi:10.1007/11569596_31

100. Rosvall M, Bergstrom CT. Maps of random walks on complex networks reveal community structure. Proc Natl Acad Sci USA. National Acad Sciences; 2008;105: 1118–1123. doi:10.1073/pnas.0706851105

101. de Castro E, Sigrist CJA, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, et al. ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. Nucleic Acids Res. 2006;34: W362–W365. doi:10.1093/nar/gkl124

102. Yang Y, Heffernan R, Paliwal K, Lyons J, Dehzangi A, Sharma A, et al. SPIDER2: A Package to Predict Secondary Structure, Accessible Surface Area, and Main-Chain Torsional Angles by Deep Neural Networks. Methods Mol Biol. New York, NY: Springer New York; 2017;1484: 55–63. doi:10.1007/978-1-4939-6406-2_6

103. McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure

prediction server. Bioinformatics. 2000;16: 404–405.

104. Drozdetskiy A, Cole C, Procter J, Barton GJ. JPred4: a protein secondary structure prediction server. Nucleic Acids Res. 2015;43: W389–W394. doi:10.1093/nar/gkv332

105. Yan R, Xu D, Yang J, Walker S, Zhang Y. A comparative assessment and analysis of 20 representative sequence alignment methods for protein structure prediction. Sci Rep. The Author(s) SN ; 2013;3: 93. doi:10.1038/srep02619

106. Cserzö M, Wallin E, Simon I, Heijne von G, Elofsson A. Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the dense alignment surface method. Protein Eng. 1997;10: 673–676.

107. Käll L, Krogh A, Sonnhammer ELL. A combined transmembrane topology and signal peptide prediction method. J Mol Biol. 2004;338: 1027–1036. doi:10.1016/j.jmb.2004.03.016

108. Rost B, Fariselli P, Casadio R. Topology prediction for helical transmembrane proteins at 86% accuracy-Topology prediction at 86% accuracy. Protein Science. Wiley-Blackwell; 1996;5: 1704–1718. doi:10.1002/pro.5560050824

109. Ikeda M, Arai M, Okuno T, Shimizu T. TMPDB: a database of experimentally-characterized transmembrane topologies. Nucleic Acids Res. Oxford University Press; 2003;31: 406–409.

110. Nugent T, Jones DT. Transmembrane protein topology prediction using support vector machines. BMC Bioinformatics. BioMed Central; 2009;10: 159. doi:10.1186/1471-2105-10-159

111. Ward JJ, McGuffin LJ, Bryson K, Buxton BF, Jones DT. The DISOPRED server for the prediction of protein disorder. Bioinformatics. 2004;20: 2138–2139. doi:10.1093/bioinformatics/bth195

112. Kozlowski LP, Bujnicki JM. MetaDisorder: a meta-server for the prediction of intrinsic disorder in proteins. BMC Bioinformatics. 2012;13: 111. doi:10.1186/1471-2105-13-111

113. Snijder B, Sacher R, Rämö P, Liberali P, Mench K, Wolfrum N, et al. Single-cell analysis of population context advances RNAi screening at multiple levels. Mol Syst Biol. EMBO Press; 2012;8: 579. doi:10.1038/msb.2012.9

114. Lin AE, Greco TM, Döhner K, Sodeik B, Cristea IM. A proteomic perspective of inbuilt viral protein regulation: pUL46 tegument protein is targeted for degradation by ICP0 during herpes simplex virus type 1 infection. Mol Cell Proteomics. 2013;12: 3237–3252. doi:10.1074/mcp.M113.030866

115. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods. Nature Publishing Group; 2012;9: 671–675. doi:10.1038/nmeth.2089

116. Cristea IM, Williams R, Chait BT, Rout MP. Fluorescent proteins as

proteomic probes. Mol Cell Proteomics. American Society for Biochemistry and Molecular Biology; 2005;4: 1933–1941. doi:10.1074/mcp.M500227-MCP200

117. Wiśniewski JR, Zougman A, Nagaraj N, Mann M. Universal sample preparation method for proteome analysis. Nat Methods. Nature Publishing Group; 2009;6: 359–362. doi:10.1038/nmeth.1322

Supporting Information

S1 Text. Details on Immunoaffinity Purification Quantitative Mass Spectrometry experiments.

S1 Fig. Modularity values of the consensus partition as a function of the filtering threshold. Modularity values achieved by each of the partitions derived from the consensus agreement matrix, by filtering values below the specified thresholds (ranging from 0 to 1 in increments of 0.5). The dashed line is only shown to emphasise the trend of the modularity values. The yellow-shaded area of the plot highlights the threshold range for which the highest modularity was achieved.

S2 Fig. Primary sequence analysis of US10. Previously reported and newly identified features are indicated. Predictions from each software tool are shown. Predicted disordered regions, α-helices and transmembrane helices are indicated in blue, yellow and pink, respectively. The identified CLRs are shown in red boxes. Individual prolines are highlighted in red. The four-residues polyproline sequence is indicated with a black box. The previously identified consensus zinc finger sequence [33] is underscored. The final assignment of the secondary structure elements was based on the consensus of individual methods (prediction confidence scores were not taken into account).

S3 Fig. Primary sequence analysis of UL55. Predictions from each software tool are shown. Predicted disordered regions, α-helices and β-strands are indicated in blue, yellow and green, respectively. The final

assignment of the secondary structure elements was based on the consensus of individual methods (prediction confidence scores were not taken into account).

S1 Table. Herpesvirus species for which PPI data were collected as input for the PPI network assembly framework.

S2 Table. Taxonomic identifiers associated to species in S1 Table and used to extract PPIs from input resources.

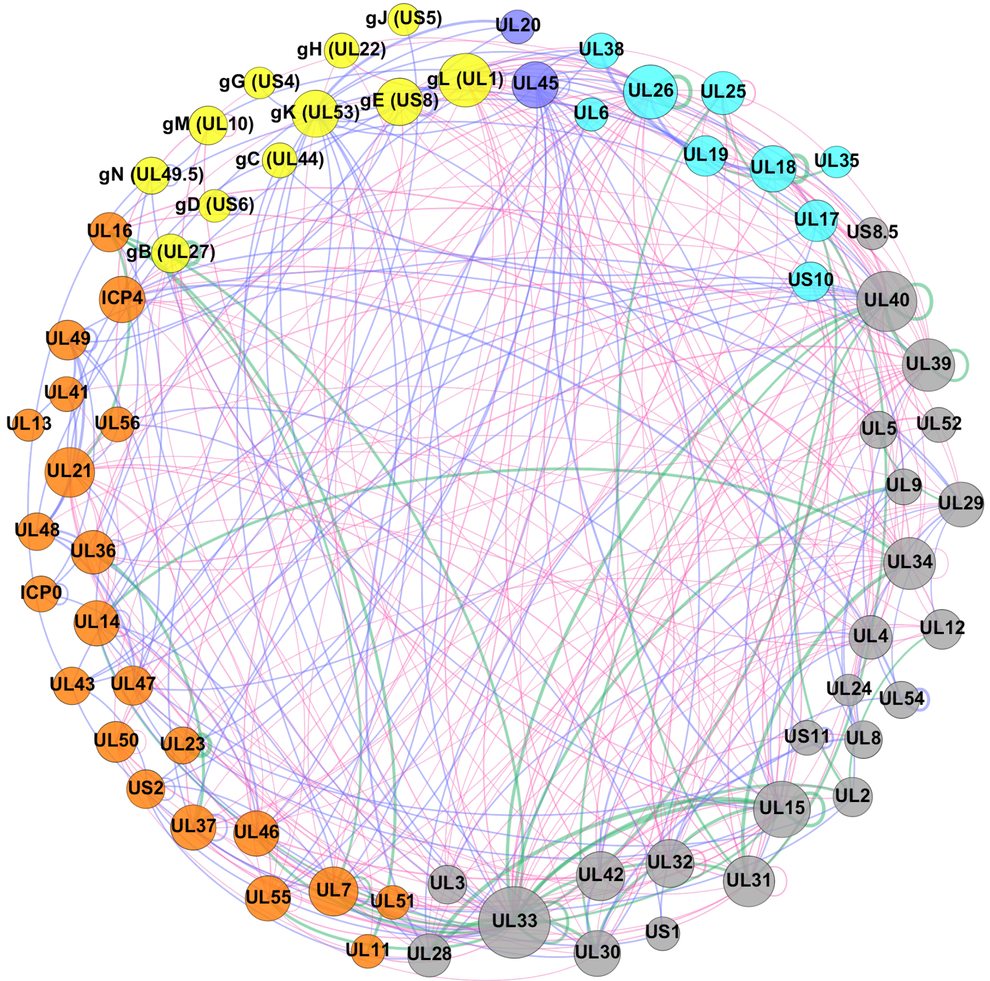
S3 Table. Protein-protein interaction network reconstructed for HSV1. For each interaction, the interacting proteins, detection methods, associated PubMed IDs, types of interaction, confidence score, and whether the interaction was computationally predicted and/or experimentally supported, is indicated.

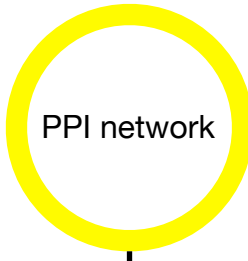
S4 Table. Parameters associated to the partitions returned by the 14 clustering algorithms.

S5 Table. Modularity values associated to the partitions derived from the consensus agreement matrix, at different thresholds.

S6 Table. Proteins co-purifying with pUL37 by immunoaffinity purification mass-spectrometry.

Figure 2.



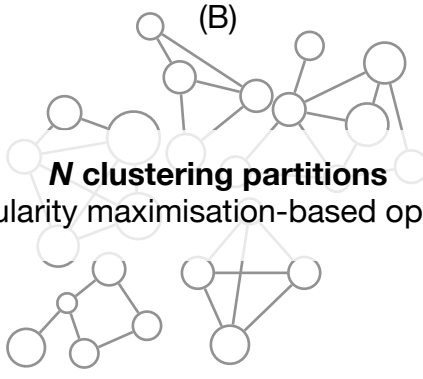


(A)

Noise reduction

- Virion components only
- Remove PPIs between capsid and envelope proteins
- Extract LCC

(B)

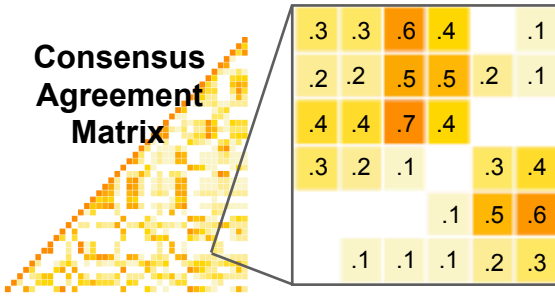


***N* clustering partitions**

- Modularity maximisation-based optimisation

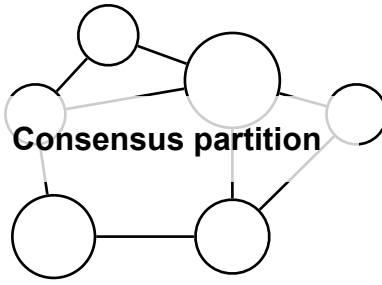
Modularity > 0

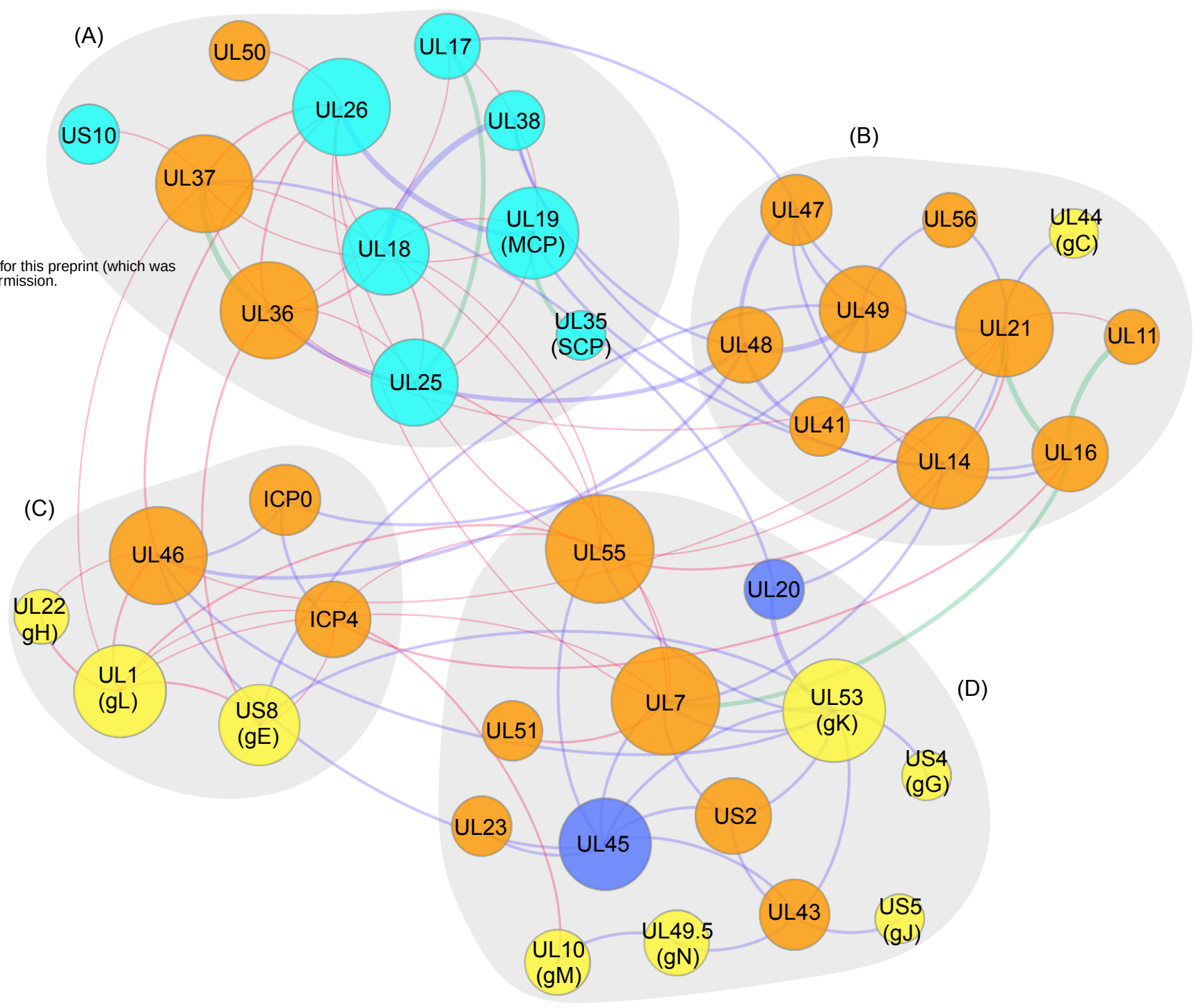
(C)



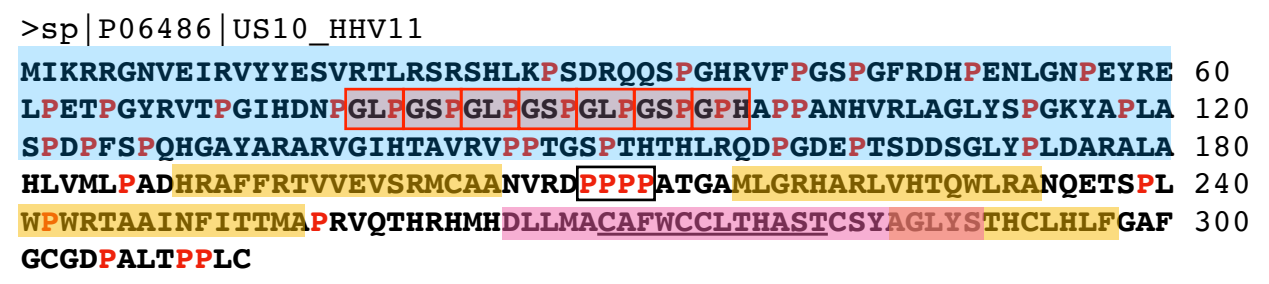
Modularity maximisation-based threshold

(D)

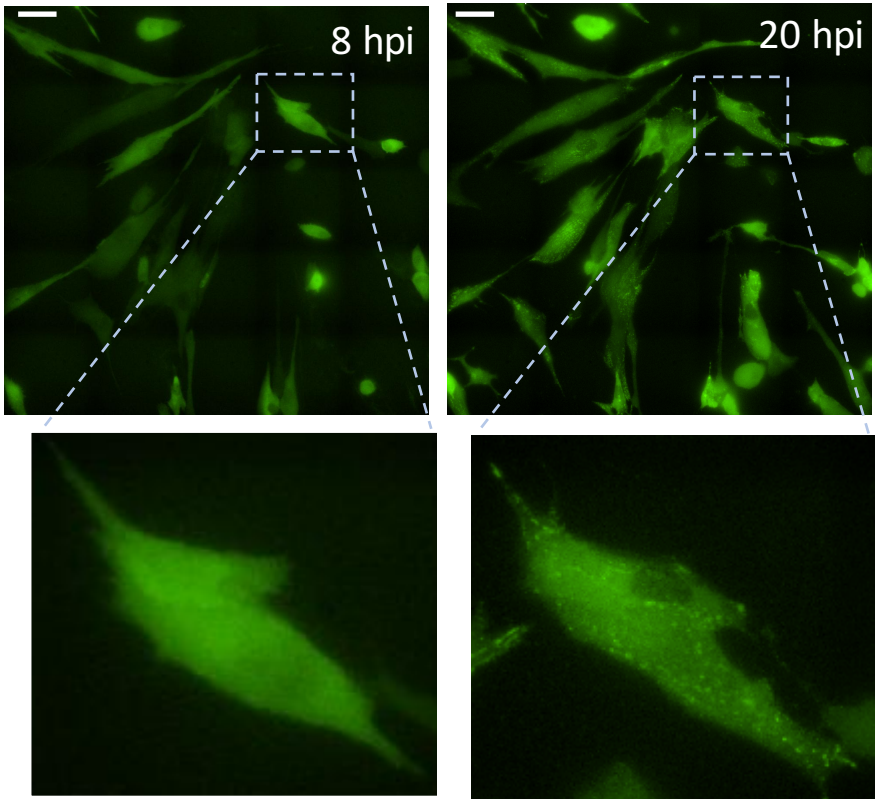




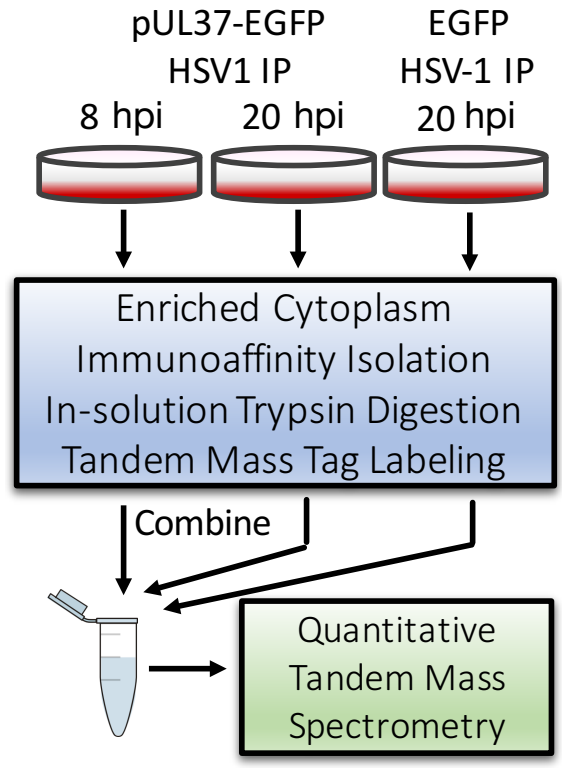
for this preprint (which was
rmission.


>sp|P06486|US10_HHV11
MIKRRGNVEIRVYYESVRTLRSRSHLKPSDRQQSPGHRVFPGSPGFRDHPENLGNPEYRE 60
LPETPGYRVTPGIHDNPGLPGSPGLPGSPGLPGSPGLPGSPGPHAPPANHVRLAGLYSPGKYAPLA 120
SPDPFSPQHGAYARARVGIHTAVRVPPTGSPTHTHLRQDPGDEPTSDDSGLYPLDARALA 180
HLVMLPADHRAFFRTVVEVSRMCAANVRDPPPPATGAMLGRHARLVHTQWLRANQETSPL 240
WPWRTAAINFITTMAPRVQTHRHMHDLLMACAFWCCLTHASTCSYAGLYSTHCLHLFGAF 300
GCGDPALTPPLC

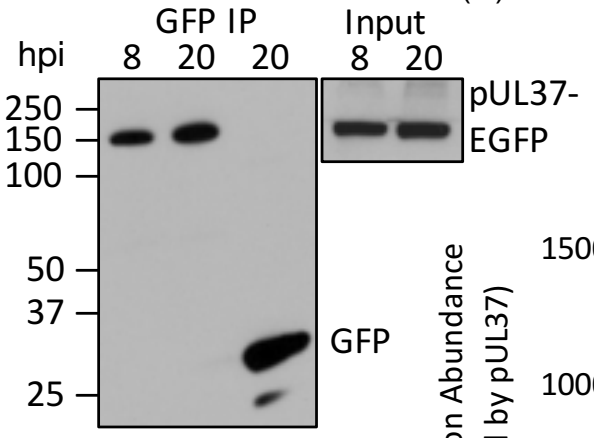
(A)

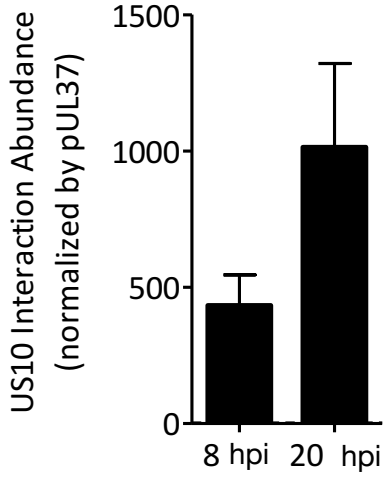

8 hpi
20 hpi

(B)

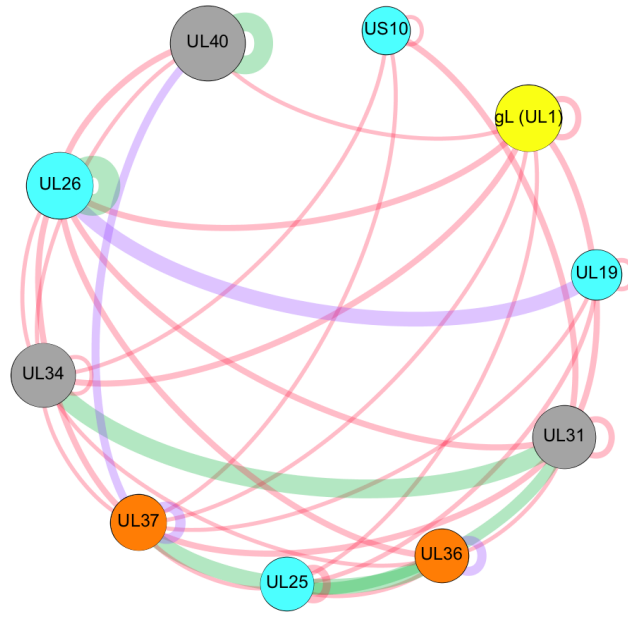

pUL37-EGFP
HSV1 IP
8 hpi
20 hpi
EGFP
HSV-1 IP
20 hpi
Enriched Cytoplasm
Immunoaffinity Isolation
In-solution Trypsin Digestion
Tandem Mass Tag Labeling
Combine
Quantitative
Tandem Mass
Spectrometry

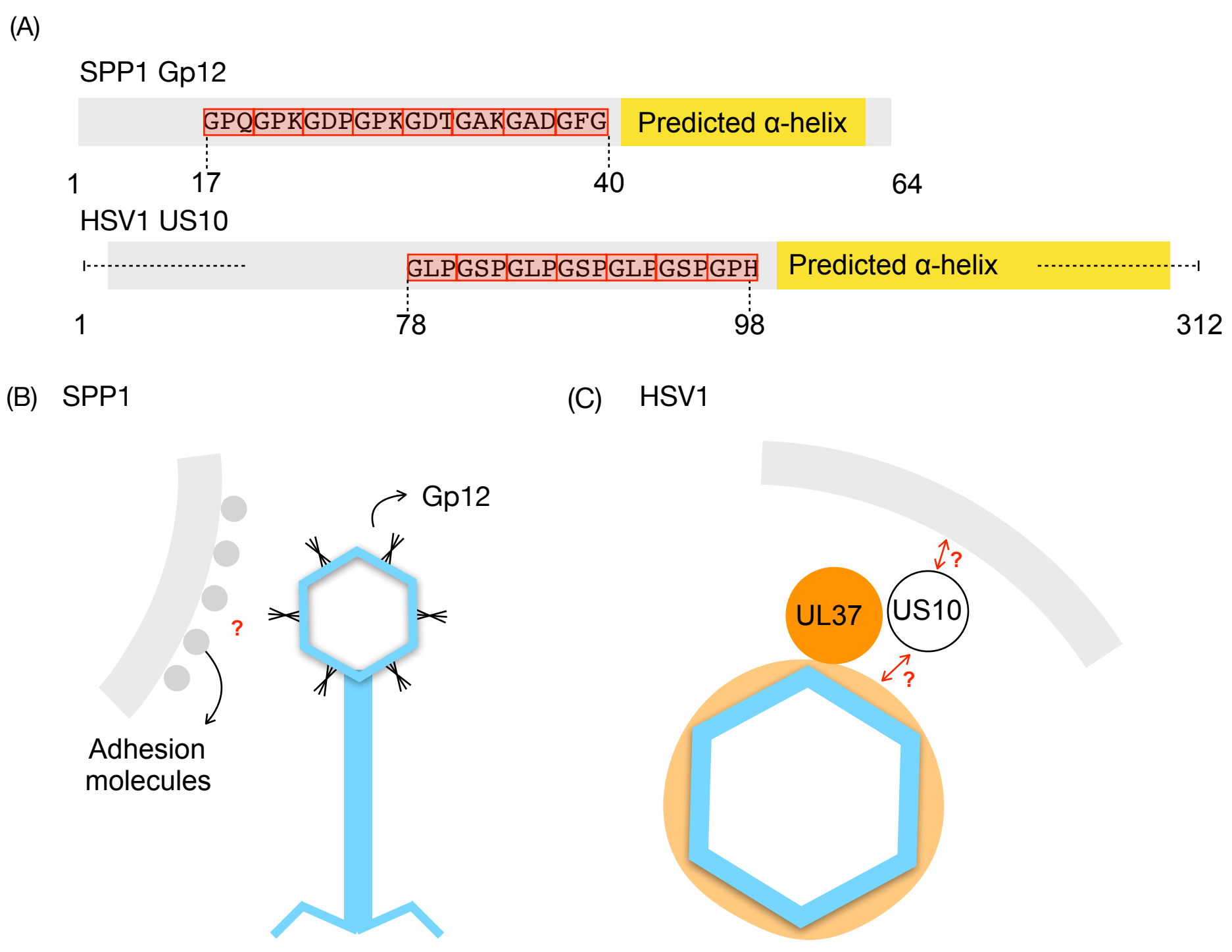
(C)

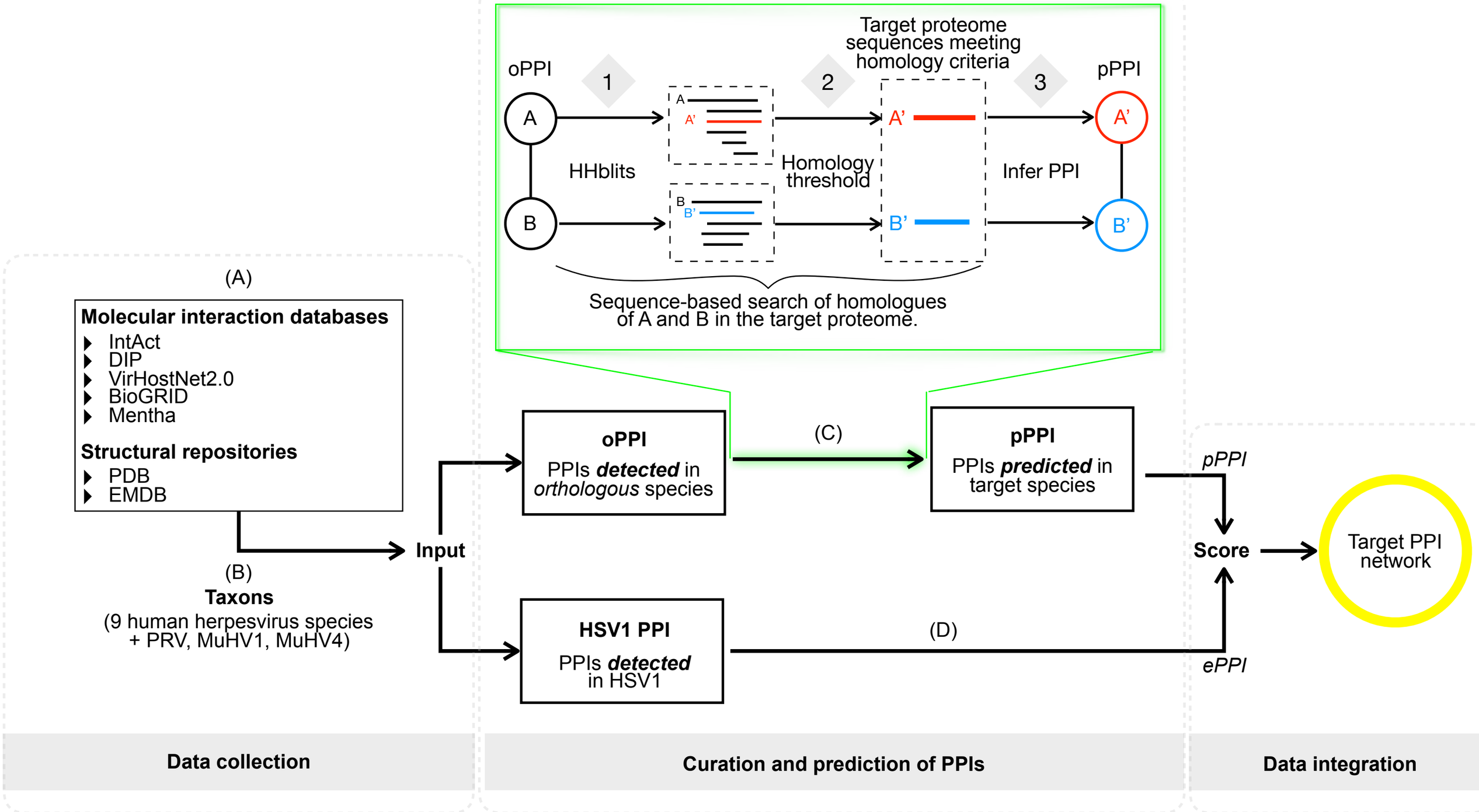

GFP IP
Input
hpi
8
20
20
8
20
250
150
100
50
37
25
pUL37-
EGFP
GFP

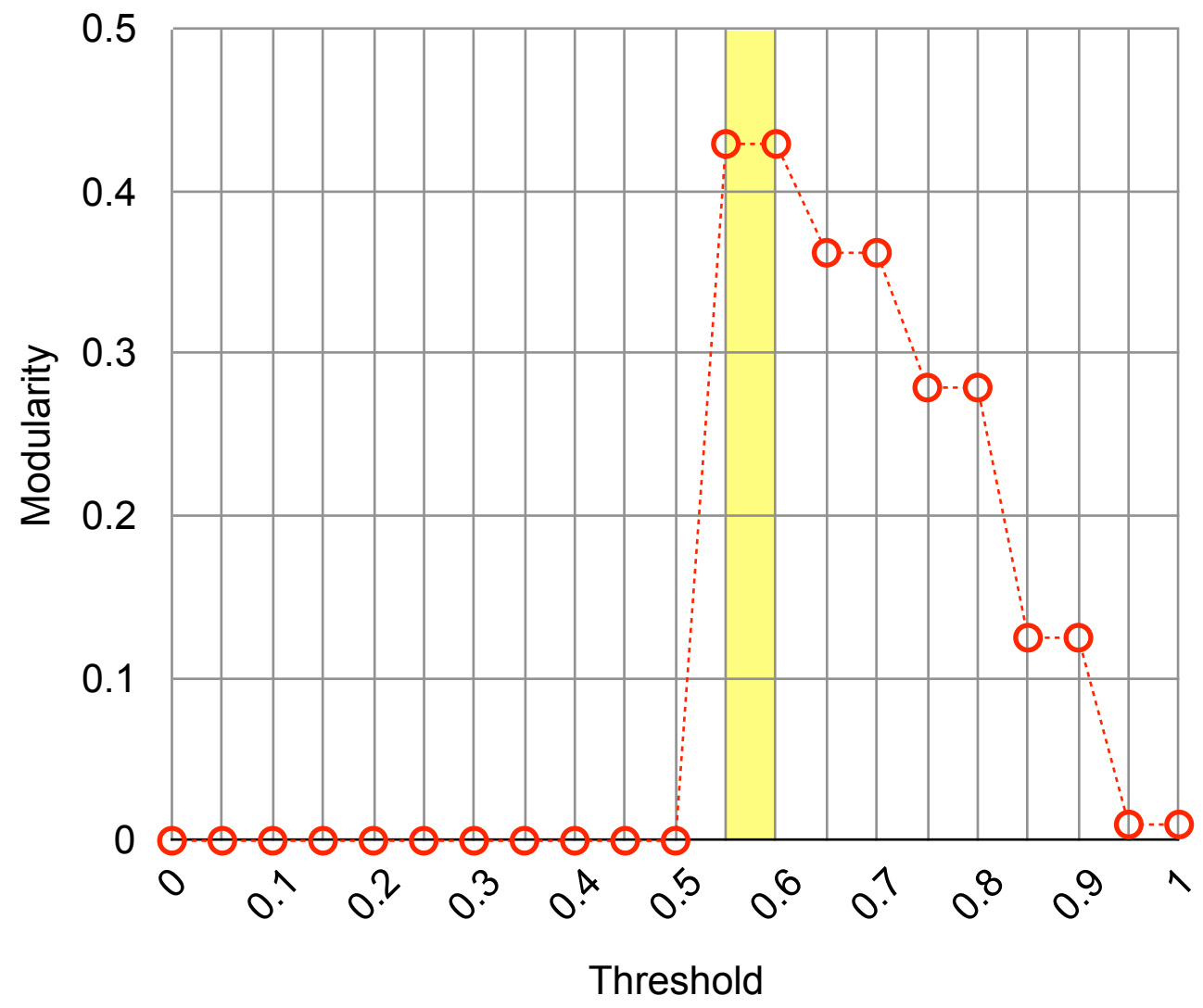

US10 Interaction Abundance
(normalized by pUL37)
1500
1000
500
0
8 hpi
20 hpi

(D)


UL40
US10
gL (UL1)
UL26
UL19
UL34
UL31
UL37
UL36
UL25


(A)
SPP1 Gp12
GPQGPKGDPGPKGDTGAKGADGFG
Predicted α-helix
1
17
40
64
HSV1 US10
GLPGSPGLPGSPGLPGSPGPH
Predicted α-helix
1
78
98
312
(B) SPP1
Gp12
?
Adhesion molecules
(C) HSV1
UL37
US10
?
?

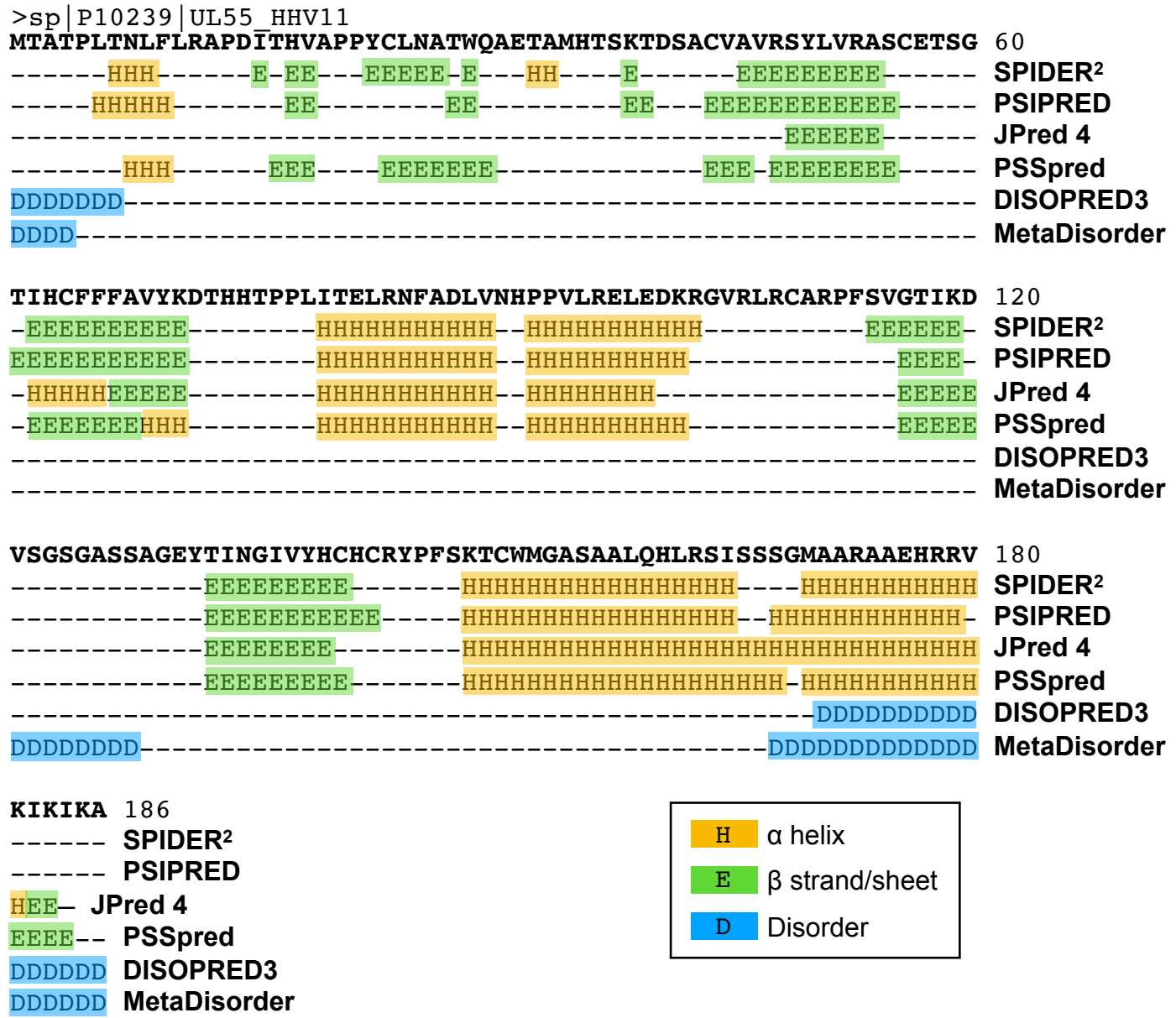

oPPI
1
2
3
pPPI
Target proteome sequences meeting homology criteria
A
A'
HHblits
Homology threshold
Infer PPI
B
B'
Sequence-based search of homologues of A and B in the target proteome.
(A)
Molecular interaction databases
IntAct
DIP
VirHostNet2.0
BioGRID
Mentha
Structural repositories
PDB
EMDB
(B)
Taxons
(9 human herpesvirus species + PRV, MuHV1, MuHV4)
Input
oPPI
PPIs detected in orthologous species
(C)
pPPI
PPIs predicted in target species
HSV1 PPI
PPIs detected in HSV1
(D)
pPPI
ePPI
Score
Target PPI network
Data collection
Curation and prediction of PPIs
Data integration


0.5
0.4
0.3
0.2
0.1
0
Modularity
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Threshold

>sp|P06486|US10_HHV11

```
MIKRRGNVEIRVYYESVRTLRSRSHLKPSDRQQSPGHRVFPGSPGFRDHPENLGNPEYRE 60
------------------------------------------------------------ MEMSAT-SVM
------------------------------------------------------------ TMpred
------------------------------------------------------------ Phobius
------------------------------------------------------------ PHDhtm
------------------------------------------------------------ DAS
--------EEEEE------------------------------------------------ SPIDER²
---------EEEE------------------------------------------------ PSIPRED
----------EEE------------------------------------------------ JPred 4
------------------------------------------------------------ PSSpred
-DDD---------------DDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDD DISOPRED3
DDDDDDD--------DDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDD MetaDisorder

LPETPGYRVTPGIHDNPGLPGSPGLPGSPGLPGSPGPHAPPANHVRLAGLYSPGKYAPLA 120
------------------------------------------------------------ MEMSAT-SVM
------------------------------------------------------------ TMpred
------------------------------------------------------------ Phobius
------------------------------------------------------------ PHDhtm
------------------------------------------------------------ DAS
------------------------------------------------------------ SPIDER²
------------------------------------------------------------ PSIPRED
------------------------------------------------------------ JPred 4
------------------------------------------------------------ PSSpred
DDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDD DISOPRED3
DDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDD MetaDisorder

SPDPFSPQHGAYARARVGIHTAVRVPPTGSPTHTHLRQDPGDEPTSDDSGLYPLDARALA 180
------------------------------------------------------------ MEMSAT-SVM
------------------------------------------------------------ TMpred
------------------------------------------------------------ Phobius
------------------------------------------------------------ PHDhtm
------------------------------------------------------------ DAS
------------------------------------------------------HHH---- SPIDER²
-------------------------------------------------------H---- PSIPRED
------------------------------------------------------------ JPred 4
------------------------------------------------------------ PSSpred
DDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDD DISOPRED3
DDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDD---------- MetaDisorder

HLVMLPADHRAFFRTVVEVSRMCAANVRDPPPPATGAMLGRHARLVHTQWLRANQETSPL 240
------------------------------------------------------------ MEMSAT-SVM
------------------------------------------------------------ TMpred
------------------------------------------------------------ Phobius
------------------------------------------------------------ PHDhtm
------------------------------------------------------------ DAS
HHHH----HHHHHHHHH-HHHHHHHH------------HHHHHHHHHHHHHHHH-------- SPIDER²
HH----------HHHHHH-HHHHH------------HHHHHHHHHHHHHHHHH-------- PSIPRED
--------------------------------------HHHHHHHH---------------- JPred 4
------------------------------------------------------------ PSSpred
DD------------------------------------------------------------ DISOPRED3
------------------------------------------------------------ MetaDisorder

WPWRTAAINFITTMAPRVQTHRHMHDLLMACAFWCCLTHASTCSYAGLYSTHCLHLFGAF 300
----------------------------TTTTTTTTTTTTTTT---------------- MEMSAT-SVM
----------------------TTTTTTTTTTTTTTTTT------------------- TMpred
----------------------TTTTTTTTTTTTTTTTTTTTTTTT------------ Phobius
----------------------------TTTTTTTTTTTTTTTTTT------------ PHDhtm
------------------------TTTTTTTTTTTTT-----------------TTT DAS
HHHHHHHHHHHHHH-------HHHHHHHHHHHHHHHHHHHH----HHHHHHHHHHHH-- SPIDER²
HH------------------------------------------------------------ PSIPRED
--HHHHHHHHHHH----------HHHHHHHHHHHH---------HHHHHHHHHHH---- JPred 4
--------------------HHHHHHHHHHHHHH----------HHHHHHHHHHH-- PSSpred
HHHHHHHHHHHHHH------HHHHHHHHHHHHHHHHHHHH-----HHHHHHHHHHHH-- DISOPRED3
------------------------------------------------------------ MetaDisorder

GCGDPALTPPLC
------------ MEMSAT-SVM
------------ TMpred
------------ Phobius
------------ PHDhtm
T----------- DAS
----------E- SPIDER²
------------ PSIPRED
------------ JPred 4
------------ PSSpred
------------ DISOPRED3
------------ MetaDisorder
```

H	α helix
E	β strand/sheet
D	Disorder
T	Transmembrane
	Collagen-like repeat
	Polyproline
	Zinc-finger consensus


>sp|P10239|UL55_HHV11
MTATPLTNLFLRAPDITHVAPPYCLNATWQAETAMHTSKTDSACVAVRSYLVRASCETSG 60
SPIDER²
PSIPRED
JPred 4
PSSpred
DISOPRED3
MetaDisorder
TIHCFFFAVYKDTHHTPPLITELRNFADLVNHPPVLRELEDKRGVRLRCARPFSVGTIKD 120
SPIDER²
PSIPRED
JPred 4
PSSpred
DISOPRED3
MetaDisorder
VSGSGASSAGEYTINGIVYHCHCRYPFSKTCWMGASAALQHLRSISSSGMAARAAEHRRV 180
SPIDER²
PSIPRED
JPred 4
PSSpred
DISOPRED3
MetaDisorder
KIKIKA 186
SPIDER²
PSIPRED
JPred 4
PSSpred
DISOPRED3
MetaDisorder
H α helix
E β strand/sheet
D Disorder