Prediction and analysis of human-herpes simplex virus type 1 protein-protein interactions by integrating multiple methods

Xianyi Lian
China Agricultural University https://orcid.org/0000-0001-7576-2198
Xiaodi Yang
China Agricultural University
Jiqi Shao
China Agricultural University
Fujun Hou
Zhejiang University
Shiping Yang
China Agricultural University
Dongli Pan
Zhejiang University
Ziding Zhang (zidingzhang@cau.edu.cn)
https://orcid.org/0000-0002-9296-571X

Research article

Keywords: Human-virus interaction, Protein-protein interaction, Prediction, Herpes simplex virus type 1, Alzheimer's disease

Posted Date: February 6th, 2020

DOI: https://doi.org/10.21203/rs.2.22765/v1

Version of Record: A version of this preprint was published at Quantitative Biology on December 1st, 2020. See the published version at https://doi.org/10.1007/s40484-020-0222-5.

Abstract

Background: Herpes simplex virus type 1 (HSV-1) is a ubiquitous infectious pathogen that widely affects human health. To decipher the complicated human-HSV-1 interactions, a comprehensive protein-protein interaction (PPI) network between human and HSV-1 is highly demanded.

Results: To complement experimental identification of human-HSV-1 PPIs, we developed an integrative strategy to predict proteome-wide PPIs between human and HSV-1. For each human-HSV-1 protein pair, four popular PPI inference methods, including interolog mapping, the domain-domain interaction-based method, domain-motif interaction-based method and machine learning-based method, were optimally implemented and employed to generate four interaction probability scores, and then these four scores were further integrated into a final probability score. As a result, a comprehensive high-confidence PPI interaction network between human and HSV-1 was established, which covers 10,432 interactions between 4,546 human proteins and 72 HSV-1 proteins. Functional and network analyses of the HSV-1 targeting proteins in the context of human interactome can recapitulate the known knowledge regarding the HSV-1 replication cycle, supporting the overall reliability of the predicted PPI network. Considering that HSV-1 infections are implicated in encephalitis and neurodegenerative diseases, we focused on exploring the biological significance of brain-specific human-HSV-1 PPIs. In particular, the predicted interactions between HSV-1 proteins and Alzheimer's-disease-related proteins were intensively investigated.

Conclusions: The current work can provide testable hypotheses to assist mechanistic understanding of the human-HSV-1 relationship as well as the discovery of anti-HSV-1 pharmaceutical targets. To make the predicted PPI network and the datasets freely accessible to the scientific community, a user-friendly database browser has been released at http://www.zzdlab.com/HintHSV/index.php.

Keywords: Human-virus interaction, Protein-protein interaction, Prediction, Herpes simplex virus type 1, Alzheimer's disease

1 Background

Herpes simplex virus type 1 (HSV-1) is a neurotropic, enveloped, and double-stranded linear DNA virus [1–4]. The genome of HSV-1 is roughly 152 kb, encoding more than 74 different genes [3]. As a widely distributed infectious virus, it can spread from person to person through direct contact. It is estimated by the World Health Organization that around 3.7 billion people under the age of 50 are infected with HSV-1 worldwide [5]. Once entering the human body from the skin or mucosa, HSV-1 can enter sensory neurons and be transported through axons to the trigeminal ganglion where it establishes a latent infection.

When stimulated, the latent virus can be reactivated to cause symptomatic or asymptomatic recurrent infections, leading to common cold sores, blisters and even other serious diseases [2–4, 6]. HSV-1 can also reach the central nervous system (CNS), occasionally leading to fatal neurological diseases such as herpes simplex encephalitis (HSE) [7, 8]. Moreover, a growing body of evidence points to a strong association between HSV-1 infection and Alzheimer's disease (AD) [9]. No existing antiviral drug can

eliminate HSV-1 infection due to the ability of the virus to undergo latent infection thereby evading drug actions. Therefore, more fundamental research efforts are required to decipher the complicated human-HSV-1 interactions, with the prospect of providing some hints for developing novel prophylactic or therapeutic methods against viral infections.

Investigations on protein-protein interactions (PPIs) between host and pathogen can reveal key biological processes in relation to host-pathogen interaction and elucidate the underlying mechanisms of infectious diseases. As an important branch of host-pathogen PPI studies, human-virus PPI has always been a focus in view of the close relationship with human diseases and health. The existing human-HSV-1 interactome researches focus on certain proteins of the virus, such as glycoproteins (involved in HSV-1 entry into host cells [10]), ICP34.5 (neurovirulence factor [11]), ICP0 (viral E3 ubiquitin ligase [12]), ICP8 (single-stranded DNA-binding protein [13]), ICP4 (Major viral transcription factor [14]) and so on. The available data are obviously not sufficient to form a PPI network that allows us to take a global view of the interactome between human and HSV-1. In general, the experimental identification of PPIs including human-virus PPIs is time-consuming, labor intensive and expensive. In this context, cost-effective computational prediction methods are playing an increasingly important role in supplementing the experimental identification of PPIs.

A plethora of host-pathogen PPI including human-virus PPI prediction methods have been developed [15–18], which were mainly borrowed from intra-species PPI prediction methods [19–21]. In principle, traditional intra-species PPI prediction methods, such as the interolog mapping (IM) [22], the domain-domain interaction (DDI)-based method [22, 23] and the domain-motif interaction (DMI)-based method [24], can be readily adapted to predict human-virus PPIs. With the accumulation of experimentally verified human-virus PPI data, machine learning (ML)-based prediction methods have been increasingly popular in the past decade. Although none of the existing human-virus PPI prediction methods can achieve satisfactory performance, it has been common knowledge that integrating multiple prediction methods can often yield more powerful and robust predictive performance, which has been implemented in a series of studies [21, 25, 26].

In this work, we integrated four PPI inference methods to predict the high-confidence PPIs between human and HSV-1 across the entire proteome (Fig. 1). In addition to the ML-based method, we also refined and incorporated three traditional PPI prediction methods (i.e. IM, DDI and DMI). Each individual prediction method yields an interaction probability score for any query protein pair, and then the four predictive scores for the query protein pair were further integrated into a final score. PPIs with higher final scores (integration score > 0.5) were singled out for further analysis. In addition to the general functional and network topology analyses of HSV-1 targeting human proteins, we further explored the biological significance of the predicted human-HSV-1 interactome by focusing on brain tissue-specific PPIs. In particular, the potential mechanisms of HSE and AD in the context of human-HSV-1 interactome were investigated.

2 Methods

2.1 Datasets

2.1.1 HSV-1 and human proteins

In this work, we focused on the PPI prediction between the HSV-1 strain KOS and human. We downloaded all the proteins of the HSV-1 strain KOS from GenBank (https://www.ncbi.nlm.nih.gov/nuccore/952947517/). By merging two redundant proteins (Protein RL2 repeats with protein RL2_1; Protein RS1 repeats with protein RS1_1; the results are presented as RL2/RL2_1 and RS1/RS1_1, respectively), we obtained 74 HSV-1 proteins (Additional file 1 Data set S1). A total number of 20,412 reviewed human proteins used for prediction was downloaded from the UniProt database [27] (Additional file 2 Data set S2).

2.1.2 Brain-specific human genes

We downloaded brain-specific genes, which reveal elevated expression in cerebral cortex, from the Human Protein Atlas (HPA, www.proteinatlas.org). By UniProt ID mapping, 1,442 brain-specific human proteins were obtained.

2.1.3 AD-related human genes

We downloaded the gene-disease associations from DisGeNET (http://www.disgenet.org/). The resulting 1,947 AD-related human genes were obtained by the UniProt ID mapping tool.

2.1.4 Human PPI network

The human interaction network was obtained from our previous work, which consists of 345,064 PPIs and 18,473 proteins. It was used for network parameter analysis and network-based encoding in the development of the ML-based predictive model. We used the R package called igraph [28] to calculate the network parameters of protein nodes in the network.

2.2 PPI prediction methods

To ensure that the predictions are robust and reliable, we used four prediction methods (IM, DDI, DMI and ML) to infer the PPIs between 74 HSV-1 proteins and 20,412 human proteins. The four methods would give the probability scores (0 ~ 1) of interaction for 74*20,412 protein pairs, respectively. Finally, the four scores would be combined into one final score according to the integration method used in the STRING database (36). Each method is briefly described as follows.

2.2.1 Interolog mapping method

The IM method is a widely used PPI inference method. The core idea of IM is to infer unknown PPIs from known homologous PPIs (termed as interologs) in other organisms. Previous IM applications often used the PPI templates from one or several model species to infer unknown PPIs. To maximize the IM method, we extended the species source range of template PPIs to cover most of the experimentally identified

PPIs, including both of intra-species and inter-species PPIs. Here, 571,359 template PPIs with relatively complete information were collected from seven public databases, including BioGRID [29], DIP [30], HPIDB [31], IntAct [32], PATRIC [33], InnateDB [34] and VirHostNet [35]. We employed the strategy of HIPPIE [36] to evaluate the quality of each PPI template. For each PPI template, a quality score (S_{temp}) ranging from 0 to 1 was assigned by accounting for three conditions (i.e. the experimental methods for the PPI determination, the literature reporting the PPI, and the species included in the PPI). To identify the interologs for a query protein pair between human and HSV-1, BLAST searching was conducted to identify their homologs, and the criteria for two proteins to be considered homologous are as follows: E-value $\leq 10^{-5}$, sequence identity $\geq 30\%$, and alignment coverage of query protein $\geq 40\%$. In case n homologous pairs were identified for the query pair, the IM-based interaction probability (Pr_{IM}) can be defined as:

$$Pr_{IM}=1-\prod_{i=1}^{n}(1-s_i),\ s_i=\begin{cases}0, \text{ if protein pair } i \text{ not in PPI templates}\\ S_{temp}, \text{ if protein pair } i \text{ in PPI templates}\end{cases} (1)$$

2.2.2 Domain-domain interaction-based method

Considering that the interaction between two proteins may be mediated through evolutionally-conserved interacting domain pairs existing in the proteins, the DDI method was developed for predicting PPIs. The list of known DDIs can be downloaded from the 3did database [37]. In order to construct a DDI library as large as possible, the expectation maximization (EM)-based algorithm proposed by Liu et al. [38] was also employed to derive domain interaction pairs from known PPIs in this study. Here, the domain definition was based on the Pfam database [39], and hmmscan [40] was employed to search protein domains (E-value $\leq 10^{-5}$). Among the known PPIs collected by us, 918,116 PPIs conform to the requirement that the corresponding two protein partners contain Pfam domains. The probability of DDIs contained in these PPIs were evaluated using the EM algorithm. Because some domains frequently occurred in proteins may not be really participated in PPIs, to avoid the introduction of potential noise, such highly frequently occurred domains were not taken into account in the subsequent implementation of the EM algorithm. With the principle that DDIs collected from 3did should be more reliable, the score of known DDIs in 3did were assigned as 1. Finally, a comprehensive DDI library was compiled by combining the known DDIs in 3did with the inferred DDIs through the EM algorithm, the confidence score (S_{DDI}) for each DDI in the library is assigned as the following formula:

$$S_{DDI}=\frac{1}{2}*(S_{DDI\text{-}EM}+S_{DDI\text{-}3did}) (2)$$

Where $S_{DDI-3did}$ takes 1 or 0 respectively to represent whether the DDI from 3did or not, and S_{DDI-EM} takes the score from the EM algorithm (0 ~ 1) or 0 (can't be inferred from EM). The probability of interaction (Pr_{DDI}) between one HSV-1 protein and one human protein is inferred from the n domain pairs they contain, which is defined as:

$$Pr_{DDI}=1-\prod_{i=1}^{n}(1-s_i),\ s_i=\begin{cases}0, \text{ if domain pair } i \text{ not in DDI library}\\ S_{DDI}, \text{ if domain pair } i \text{ in DDI library}\end{cases} \quad (3)$$

2.2.3 Domain-motif interaction-based method

Domain-motif interaction is also an important way to mediate human-virus PPIs. Like the DDI method, the DMI method can also be employed to infer PPIs. The DMI library is also a combination of known DMIs and the inferred DMIs with the assistance of the EM algorithm. Known DMIs were also downloaded from 3did. Here domain assignment is the same as the DDI method. We only focused on those motif patterns in known DMIs and obtained each protein and its motifs by regular expression matching. Similar to the filtering strategy used in the DDI method, we removed the evaluated DMIs containing the highly frequently occurred domains or motifs before scoring them with the EM algorithm. Finally, the confidence score (S_{DMI}) for each DMI in the library is defined using the following equation:

$$S_{DMI}=\frac{1}{2}*(S_{DMI-EM}+S_{DMI-3did}) \quad (4)$$

Where $S_{DMI-3did}$ takes 1 or 0 respectively to represent whether the DMI from 3did or not, and S_{DMI-EM} is the score of the DMI from the EM algorithm (0 ~ 1) or 0 (can't infer from EM). The interaction probability (Pr_{DMI}) of a human-HSV-1 protein pair containing n domain-motif pairs is further inferred from the following formula:

$$Pr_{DMI}=1-\prod_{i=1}^{n}(1-s_i),\ s_i=\begin{cases}0, \text{ if domain-motif pair } i \text{ not in DMI library}\\ S_{DMI}, \text{if domain-motif pair } i \text{ in DMI library}\end{cases} \quad (5)$$

2.2.4 Machine learning-based method

To develop an ML-based predictor, we compiled a training dataset containing 728 positive samples (i.e. known human-HSV-1 PPIs) and 7,280 negative samples (non-PPIs). The positive samples were collected from HPIDB, in which HSV-1 proteins from different strains were considered, while the negative samples

were randomly selected from human-HSV-1 protein pairs with unidentified interaction relationships. Moreover, two encoding schemes were employed to transform protein pairs into feature vectors, including a sequence-based encoding called CKSAAP and a network property-based encoding called NetTP. More details about these two encoding schemes are available in our previous publication [41]. Subsequently, the predictive models of the two encoding methods were trained respectively by the random forest method, and then they were integrated into a stronger predictive model through logistic regression. The performance of the two individual models and integrative model was evaluated through 5-fold cross-validation (Additional file 4 Fig. S1). In general, the integrative model can outperform each individual ML model. For each query protein pair, the final prediction model would generate a prediction score (S_{ML}) ranging from 0 to 1. We calculated the F1 values of the model in the 5-fold cross-validation according to different thresholds and took the threshold value of 0.363 corresponding to the maximum value of F1 as the final criterion to determine whether the query pair have interaction or not. The formula of F1 value can be found in our previously publication [41]. Furthermore, the prediction score was converted into the ML-based interaction probability score (Pr_{ML}):

$$\mathrm{Pr_{ML}} = \begin{cases} \mathrm{S_{ML}},\ \mathrm{S_{ML}} \geq 0.363 \\ 0,\ \mathrm{S_{ML}} < 0.363 \end{cases} \quad (6)$$

2.3 ID mapping

The online UniProt ID mapping tool (https://www.uniprot.org/uploadlists/) was used to convert other IDs such as human or viral gene IDs into UniProt IDs.

2.4 Gene Ontology enrichment analysis

We used the BiNGO plugin [42] in Cytoscape [43] for Gene Ontology (GO) enrichment analysis. The enrichment analysis of UL22-targeted human proteins was conducted against the background of 20,412 reviewed human proteins, and the GO category of cellular component (CC) was selected. To explore why HSV-1 targets the human proteins specifically expressed in brain tissues, GO enrichment analysis of the three categories [biological process (BP), CC and molecular function (MF)] was conducted by taking the 1,442 brain-specific human proteins as the background (reference set). Statistical significance was inferred from the hypergeometric test and enriched terms were selected with a significance level of 0.05 after the Benjamini and Hochberg False Discovery Rate correction.

3 Results And Discussion

3.1 The landscape of predicted human-HSV-1 PPIs

In this work, we applied an integrative computational framework to predict the interactions between 74 different proteins of HSV-1 strain KOS and 20,412 reviewed human proteins. Our computational framework used four methods (IM, DDI, DMI and ML) to predict whether two proteins interact. Then, the four interaction probability scores (Pr_{IM}, Pr_{DDI}, Pr_{DMI}, and Pr_{ML}) were combined into an integration score (Pr) representing the interaction probability of the human-HSV-1 protein pair.

We separately calculated the number of PPIs predicted by each individual method. As shown in Fig. 2A, the number of PPIs predicted by DMI was the largest (41,828), followed by DDI (13,579), IM (7,805), and ML (6,341). In general, the percentages of overlapping PPIs among different methods are low, implying that different methods are distinctive and complementary. Due to the methodology similarity, DDI achieved a relatively more consistent PPI prediction results with IM and DMI (the overlap rate in both cases accounts for about 10% of its total). After integrating the results of the four methods, the number of predicted PPIs with $Pr > 0$ was 65,673. Although higher Pr should correspond to higher reliability, it is still necessary to set a reasonable and convincing threshold for high-confidence predictions. We sought the solution from high-throughput human-virus PPI identification studies. Taking the number of experimentally validated PPIs between HIV-1 and human as a reference, each HIV-1 protein was identified to have 100 ~ 200 interactions with human proteins in some high-throughput experimental studies [44]. Thus, we set $Pr > 0.5$ as the threshold of high-confidence PPIs (Additional file 4 Fig. S2), and 10,432 PPIs were singled out as the most likely interacting protein pairs. We found that 690 of 728 experimentally verified PPIs (collected from HPIDB database and used in the ML method) overlap with our 10,432 predicted results (Additional file 3 Data set S3) and 601 of these 690 PPIs can be predicted by more than one method. Figure 2B showed that the IM method accounts for the largest proportion among these 10,432 high-confidence PPIs.

3.2 Functional and network analysis showing the reliability of predicted human-HSV-1 PPIs

We further analyzed these 10,432 high-confidence PPIs. First, we counted the number of human proteins targeted by each HSV-1 protein (Fig. 3). On average, one HSV-1 protein interacts with 145 human proteins and the top ten HSV-1 proteins contribute 5,963 interactions (approximately 57%) in the predicted human-HSV-1 interactome. We found that the HSV-1 protein UL22 was predicted to have the most interactions with human proteins, and the predicted interaction partners are significantly enriched in the category of membrane-bounded organelle components (hypergeometric test, corrected P value = 3.37×10^{-51}). Previous studies have suggested that UL22, also called envelope glycoprotein H (gH), complexed with glycoprotein L (gL, UL1) and interacted with glycoproteins B (gB, UL27) and D (gD, US6) to form a viral membrane fusion machine, thereby driving the fusion of the virus with the host membranes to allow the virus to enter or spread between host cells [45]. It is therefore reasonable to predict that this viral protein to interact with multiple human proteins especially membrane proteins. RL2/RL2_1, E3 ubiquitin ligase (ICP0), was predicted to interact with many human proteins belonging to the category of host cellular

interferon-related proteins (hypergeometric test, corrected P value = 1.32×10^{-9}), which may indicate that RL2/RL2_1 is an HSV-1's weapon to counteract the intrinsic- and interferon-based antiviral responses. Thus, the predicted viral targets play an important role in viral infection process, which indicates the reliability of our human-virus PPI prediction.

To hijack and utilize host cells to complete viral life cycles, viral proteins tend to target some important host (human) proteins, such as the "hub" (high-degree centrality) and "bottleneck" (high-betweenness centrality) nodes of the human PPI network. Therefore, we also calculated the degree centrality and betweenness centrality of target proteins (proteins in the human PPI network that are targeted by HSV-1) and non-target proteins (proteins in the human PPI network that are not targeted by HSV-1) from the perspective of network biology. It can be clearly seen from Fig. 4 that, whether in degree or betweenness centrality, the values of target proteins were significantly higher than that of non-target proteins (Wilcoxon rank sum test, P value < 2.2×10^{-16}), which is in accordance with previous observations inferred from human-pathogen PPI network analyses.

3.3 Functional analysis of brain-specific human-HSV-1 PPIs

Among several diseases caused by the infection of HSV-1, sporadic but often fatal HSE is of great concern, which is caused by HSV-1 infection in the brain. Therefore, we further paid our attention to the PPIs in which the human proteins are specifically expressed in the brain tissue. We selected 569 PPIs containing 283 brain-specific human proteins from the 10,432 high-confidence PPIs. According to the GO enrichment analysis (Fig. 5), we found that the cell adhesion-related BP terms such as "cell adhesion", "biological adhesion" and "cell-cell adhesion" were significantly enriched (Fig. 5A, corrected P value = 2.33×10^{-7}, 2.33×10^{-7} and 2.2×10^{-14}, respectively), which indicated that HSV-1 relies on the intricate events of attachment and fusion to enter cells, specifically by utilizing its own envelope proteins (envelope glycoproteins) to interact with cell adhesion molecules to mediate this process [46]. In our results, 55 cellular adhesion molecules were predicted to interact with HSV-1 proteins. In the category of CC, we found that human proteins were significantly enriched in microtubule or microtubule cytoskeleton (Fig. 5B, corrected P value = 1.14×10^{-4} and 4.82×10^{-4}, respectively). As we know, microtubules are major components of the cytoskeleton and are involved in transport in all eukaryotic cells. Therefore, the above enriched GO terms are in accordance with previous knowledge that after entering the host cell, the viral capsids need to be transported to and from the nucleus to complete the replication cycle. This is particularly relevant to the processes of establishment of latent infection and reactivation in neurons during which transport of capsids along microtubules in long axons is required.

In addition, one of the strategies usurped by HSV-1 is to guide the entry pathway by manipulating the cell signaling cascades [47]. We found that in the GO enrichment analysis results of MF entries (Fig. 5C), the GO term of "calcium ion binding" was significantly enriched. Ca^{2+} is one of the most prominent and common signal carriers and is known to modulate several steps during virus replication. The entry of HSV-1 is triggered by the interaction of gH with cellular integrin, which eventually triggers Ca^{2+}-mediated

signaling pathways within the cell to ensure effective nucleocapsid translocation into the cytoplasm [47]. Although the relationship between chloride channels and viral infections has received less attention, previous studies have shown that chloride channels play an important role in HSV-1 entry [48]. Here we also found the significant CC enrichment of the chloride channel complex and the MF enrichment of the chloride channel activity, which further supports the association between chloride channel and HSV-1 entry.

Collectively, the GO enrichment results of HSV-1-interacting human brain-specific proteins are consistent with known functions associated with the HSV-1 replication cycle, suggesting that the PPIs between HSV-1 and human disrupt proteins normal functions in brain cells, which may cause inflammation and damage leading to HSE. These data also support the overall reliability of the predicted PPIs. We expect that the 569 PPIs can form a vital subnetwork (Additional file 4 Fig. S3) of the human-HSV-1 interactome, which may enhance mechanistic understanding of diseases related to HSV-1 infection (e.g. HSE) as well as providing new hints to therapeutic target discovery.

3.4 The association of HSV-1 with AD in the context of human-HSV-1 PPIs

Increasing evidence points to the association of HSV-1 brain infection with AD. HSV-1 is present in the latent state in a high proportion of elderly brains. Intermittent reactivation from the latent state may cause local damage and inflammation, accumulation of which might eventually lead to AD [7]. In order to investigate whether the emergence of AD is related to the HSV-1 infection, we compared 1,947 AD-related human genes with our 4,546 predicted HSV-1 target proteins (human proteins present in the 10,432 high-confidence PPIs), and 635 were found to be overlapping (Fig. 6A, hypergeometric test, P value = 1.10×10^{-28}). Meanwhile, we calculated the overlap between AD-related genes and target proteins specifically expressed in brain tissue and found that the overlap was still significant (hypergeometric test, P value = 9.18×10^{-10}). We also calculated the average network distance of AD-related genes to target proteins and non-target proteins in the human PPI network, and the results showed that AD-related genes were closer to target proteins (Fig. 6B). The above network analyses suggest that, to a large extent, HSV-1 target proteins are heavily associated with AD, and it can be hypothesized that the virus may also indirectly affect these AD-related genes by interacting with other proteins to enhance their ability to influence AD risk and predisposition.

Among our predicted PPIs, three HSV-1 proteins (UL2, UL21, and UL45) interact with amyloid precursor protein (APP). APP is a single-pass transmembrane protein that is widely expressed in tissues, especially at high levels in brain neurons and then rapidly metabolized [49]. There are two pathways for the proteolysis of APP, one of which is cleaved by α-secretase generating the sAPPα fragment and the other is cleaved by β-secretase (BACE1) producing neurotoxic amyloid β (Aβ) [49]. One of the commonly recognized hallmarks of AD is the accumulation of Aβ. Our results may raise the possibility that HSV-1 infection appears to contribute to the Aβ deposition process through PPIs (Fig. 6C). First of all, two of our

three predicted interactions are consistent with the experimental observations that HSV-1 uses its capsid proteins (UL21 and UL45) to physically interact with APP, thereby hijacking APP to transport newly generated virions in infected cells through a rapid anterograde transport mechanism [2]. Although such behavior changes the intracellular distribution of APP and seems to prevent APP from converting to Aβ to some extent, HSV-1 infection triggers an intra-CNS anti-microbial innate immune response to induce APP phosphorylation and activates BACE1 activity, which jointly promote the production of Aβ [50]. Aβ would encapsulate the HSV-1 virions to facilitate their clearance by autophagy [51, 52]. HSV-1 also employs its virulence factors (RL1 and UL45 were predicted to play a role) to counterattack, inhibiting the autophagy-lysosome pathway of Aβ by interacting with Beclin-1 [11]. The imbalance between the production and elimination of Aβ caused by HSV-1 infection accounted for excessive intracellular neurotoxic Aβ deposition within autophagosomes and endosomes, therefore inducing neuronal apoptosis, which in turn can drive degeneration of CNS tissue and development of AD. In summary, the recapitulated interactions among HSV-1, APP and Aβ argue for a mechanistic basis for the association between HSV1 infection and the risk of AD.

3.5 Interactive web interface

We stored our predicted 10,432 high confidence PPIs in a database and provided an interactive web interface (http://www.zzdlab.com/HintHSV/index.php) to facilitate user access. We have provided a search box for 72 HSV-1 proteins participating in these 10,432 PPIs, so users can select any protein to view the corresponding interactions. For each HSV-1 protein, we provide a table to display all the prediction scores for each human target protein (including four individual prediction scores and one integrative score) and a subnetwork to show the PPIs, which are available for download. The 569 brain-specific PPIs, 690 known PPIs and other datasets we used in this work are also downloadable in the web interface.

4 Conclusions

In this work, we used four popular PPI inference methods to predict the interactions between human and HSV-1. To maximize the reliability of predictions, we integrated the interaction probability scores from the four methods into a final probability score and selected a stringent threshold ($Pr > 0.5$) to single out high-confidence PPIs. The subsequent functional and network topology analysis also proved that our prediction strategy has an overall reasonable reliability in methodology. To investigate the associations between HSV-1 infection and neurodegenerative diseases (e.g. HSE and AD), we focused on brain-specific PPIs between human and HSV-1 and a subnetwork containing 569 inter-species PPIs was established. Functional analysis shows that human proteins involved in the entry, intracellular transport pathways and various regulatory pathways are utilized or hijacked by HSV-1 through complicated inter-species PPIs. Collectively, the established human-HSV-1 PPI network provides a global landscape regarding the human-HSV-1 interactome, as well as new insights into pathogenesis of HSV-1 infection.

Abbreviations

AD: Alzheimer's disease; APP: amyloid precursor protein; Aβ: amyloid β; BP: biological process; CC: cellular component; CNS: central nervous system; DDI: domain-domain interaction; DMI: domain-motif interaction; GO: Gene Ontology; HSE: herpes simplex encephalitis; HSV-1: Herpes simplex virus type 1; IM: interolog mapping; ML: machine learning; MF: molecular function; PPI: protein-protein interaction

Additional Files

Additional files

Additional file 1
Data set S1 | Information of 74 HSV-1 strain KOS proteins. (XLSX 40 kb)
Additional file 2
Data set S2 | Sequences of 20,412 reviewed human proteins with UniProt ID. (XLSX 6,969 kb)
Additional file 3
Data set S3 | 10,432 predicted high-confidence PPIs including 690 known PPIs. (XLSX 501 kb)
Additional file 4
Supplemental Fig. S1 | Precision-recall (PR) curves (A) and receiver operating characteristic (ROC) curves (B) of two individual ML models and the integrative ML model. Supplemental Fig. S2 | Number of predicted PPIs under different thresholds. Supplemental Fig. S3 | Brain-specific PPI subnetwork between human and HSV-1. (PDF 1,469 kb)

Declarations

Funding

This work was supported by the National Key Research and Development Program of China (2017YFC1200205 to Z.Z. and 2017YFC1200204 to D.P.).

Availability of data and materials

The web interface is http://www.zzdlab.com/HintHSV/index.php. The datasets used in our article can be downloaded at the web interface or according to the references mentioned in the main text.

Authors' contributions

All authors listed have made substantial, direct and intellectual contributions to the work, and approved it for publication.

Competing interests

The authors declare that there exist no conflicts of interest with regard to the present study.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

[1]State Key Laboratory of Agrobiotechnology, College of Biological Sciences, China Agricultural University, Beijing 100193, China

[2]National Demonstration Center for Experimental Biological Sciences Education, College of Biological Sciences, China Agricultural University, Beijing 100193, China

[3]Department of Medical Microbiology and Parasitology, and Department of Infectious Diseases of Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang, 310058, China

[4]State Key Laboratory of Plant Physiology and Biochemistry, College of Biological Sciences, China Agricultural University, Beijing 100193, China.

References

1. Conrady CD, Drevets DA, Carr DJJ. Herpes simplex type I (HSV-1) infection of the nervous system: is an immune response a good thing? J Neuroimmunol. 2010;220:1–9.
2. Piacentini R, Chiara G De, Domenica DLP, Ripoli C, Marcocci ME, Garaci E, et al. HSV-1 and Alzheimer's disease: more than a hypothesis. Front Pharmacol. 2014;5 May:97.
3. Watanabe D. Medical application of herpes simplex virus. J Dermatol Sci. 2010;57:75–82.

4. De Chiara G, Piacentini R, Fabiani M, Mastrodonato A, Marcocci ME, Limongi D, et al. Recurrent herpes simplex virus-1 infection induces hallmarks of neurodegeneration and cognitive deficits in mice. PLOS Pathog. 2019;15:e1007617.
5. Looker KJ, Magaret AS, May MT, Turner KME, Vickerman P, Gottlieb SL, et al. Global and regional estimates of prevalent and incident herpes simplex virus type 1 infections in 2012. PLoS One. 2015;10:e0140765.
6. 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. 2016;15:2939–53.
7. Itzhaki RF. Corroboration of a major role for herpes simplex virus type 1 in Alzheimer's Disease. Front Aging Neurosci. 2018;10:324.
8. Steiner I. Herpes simplex virus encephalitis: new infection or reactivation? Curr Opin Neurol. 2011;24:268–74.
9. Itzhaki RF, Lin W, Shang D, Wilcock GK, Faragher B, Jamieson GA. Herpes simplex virus type 1 in brain and risk of Alzheimer's disease. Lancet. 1997;349:241–4.
10. Agelidis AM, Shukla D. Cell entry mechanisms of HSV: what we have learned in recent years. Future Virol. 2015;10:1145–54.
11. Orvedahl A, Alexander D, Tallóczy Z, Sun Q, Wei Y, Zhang W, et al. HSV-1 ICP34.5 confers neurovirulence by targeting the Beclin 1 autophagy protein. Cell Host Microbe. 2007;1:23–35.
12. Smith MC, Boutell C, Davido DJ. HSV-1 ICP0: paving the way for viral replication. Future Virol. 2011;6:421–9.
13. Bryant KF, Yan Z, Dreyfus DH, Knipe DM. Identification of a divalent metal cation binding site in herpes simplex virus 1 (HSV-1) ICP8 required for HSV replication. J Virol. 2012;86:6825–34.
14. Dremel SE, Deluca NA. Herpes simplex viral nucleoprotein creates a competitive transcriptional environment facilitating robust viral transcription and host shut off. Elife. 2019;8:e51109.
15. Yang S, Fu C, Lian X, Dong X, Zhang Z. Understanding human-virus protein-protein interactions using a human protein complex-based analysis framework. mSystems. 2019;4:e00303-18.
16. Nourani E, Khunjush F, Durmuş S. Computational prediction of virus–human protein–protein interactions using embedding kernelized heterogeneous data. Mol Biosyst. 2016;12:1976–86.
17. Eid F-E, ElHefnawi M, Heath LS. DeNovo: virus-host sequence-based protein-protein interaction prediction. Bioinformatics. 2016;32:1144–50.
18. Qi Y, Tastan O, Carbonell JG, Klein-Seetharaman J, Weston J. Semi-supervised multi-task learning for predicting interactions between HIV-1 and human proteins. Bioinformatics. 2010;26:i645–52.
19. Zhou Y, Zhou YS, He F, Song J, Zhang Z. Can simple codon pair usage predict protein-protein interaction? Mol Biosyst. 2012;8:1396–404.
20. Guo Y, Yu L, Wen Z, Li M. Using support vector machine combined with auto covariance to predict protein-protein interactions from protein sequences. Nucleic Acids Res. 2008;36:3025–30.

21. Kotlyar M, Pastrello C, Pivetta F, Lo Sardo A, Cumbaa C, Li H, et al. In silico prediction of physical protein interactions and characterization of interactome orphans. Nat Methods. 2014;12:79–84.
22. Li Z-G, He F, Zhang Z, Peng Y-L. Prediction of protein-protein interactions between *Ralstonia solanacearum* and *Arabidopsis thaliana*. Amino Acids. 2012;42:2363–71.
23. Schleker S, Garcia-Garcia J, Klein-Seetharaman J, Oliva B. Prediction and comparison of *Salmonella*-human and *Salmonella-Arabidopsis* interactomes. Chem Biodivers. 2012;9:991–1018.
24. Evans P, Dampier W, Ungar L, Tozeren A. Prediction of HIV-1 virus-host protein interactions using virus and host sequence motifs. BMC Med Genomics. 2009;2:27.
25. Yang S, Li H, He H, Zhou Y, Zhang Z. Critical assessment and performance improvement of plant-pathogen protein-protein interaction prediction methods. Brief Bioinform. 2019;20:274–87.
26. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res. 2017;45:D362–8.
27. Consortium U. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res. 2019;47:D506–15.
28. Csárdi G, Nepusz T. The igraph software package for complex network research. InterJournal Complex Syst. 2006;1695:1–9.
29. 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.
30. Xenarios I, Rice DW, Salwinski L, Baron MK, Marcotte EM, Eisenberg D. DIP: the Database of Interacting Proteins. Nucleic Acids Res. 2000;28:289–91.
31. Ammari MG, Gresham CR, McCarthy FM, Nanduri B. HPIDB 2.0: a curated database for host–pathogen interactions. Database. 2016;2016:baw103.
32. Hermjakob H, Montecchi-Palazzi L, Lewington C, Mudali S, Kerrien S, Orchard S, et al. IntAct: an open source molecular interaction database. Nucleic Acids Res. 2004;32:D452–5.
33. Wattam AR, Davis JJ, Assaf R, Boisvert S, Brettin T, Bun C, et al. Improvements to PATRIC, the all-bacterial bioinformatics database and analysis resource center. Nucleic Acids Res. 2017;45:D535–42.
34. Breuer K, Foroushani AK, Laird MR, Chen C, Sribnaia A, Lo R, et al. InnateDB: systems biology of innate immunity and beyond—recent updates and continuing curation. Nucleic Acids Res. 2013;41:D1228–33.
35. 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.
36. Schaefer MH, Fontaine JF, Vinayagam A, Porras P, Wanker EE, Andrade-Navarro MA. HIPPIE: integrating protein interaction networks with experiment based quality scores. PLoS One. 2012;7:e31826.
37. Stein A, Céol A, Aloy P. 3did: identification and classification of domain-based interactions of known three-dimensional structure. Nucleic Acids Res. 2011;39:D718–23.

38. Liu X, Huang Y, Liang J, Zhang S, Li Y, Wang J, et al. Computational prediction of protein interactions related to the invasion of erythrocytes by malarial parasites. BMC Bioinformatics. 2014;15:393.
39. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, et al. The Pfam protein families database in 2019. Nucleic Acids Res. 2019;47:D427–32.
40. Eddy SR. Profile hidden Markov models. Bioinformatics. 1998;14:755–63.
41. Lian X, Yang S, Li H, Fu C, Zhang Z. Machine-learning-based predictor of human-bacteria protein-protein interactions by incorporating comprehensive host-network properties. J Proteome Res. 2019;18:2195–205.
42. Maere S, Heymans K, Kuiper M. BiNGO: a Cytoscape plugin to assess overrepresentation of Gene Ontology categories in Biological Networks. Bioinformatics. 2005;21:3448–9.
43. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13:2498–504.
44. Bandyopadhyay S, Ray S, Mukhopadhyay A, Maulik U. A review of in silico approaches for analysis and prediction of HIV-1-human protein-protein interactions. Brief Bioinform. 2014;16:830–51.
45. Szpara ML, Gatherer D, Ochoa A, Greenbaum B, Dolan A, Bowden RJ, et al. Evolution and diversity in human herpes simplex virus genomes. J Virol. 2014;88:1209–27.
46. Zhang N, Yan J, Lu G, Guo Z, Fan Z, Wang J, et al. Binding of herpes simplex virus glycoprotein D to nectin-1 exploits host cell adhesion. Nat Commun. 2011;2:510–77.
47. Azab W, Gramatica A, Herrmann A, Osterrieder N. Binding of alphaherpesvirus glycoprotein H to surface α4β1-integrins activates calcium-signaling pathways and induces phosphatidylserine exposure on the plasma membrane. MBio. 2015;6:e01552-15.
48. Zheng K, Chen M, Xiang Y, Ma K, Jin F, Wang X, et al. Inhibition of herpes simplex virus type 1 entry by chloride channel inhibitors tamoxifen and NPPB. Biochem Biophys Res Commun. 2014;446:990–6.
49. O'Brien RJ, Wong PC. Amyloid precursor protein processing and Alzheimer's disease. Annu Rev Neurosci. 2011;34:185–204.
50. Hogestyn JM, Mock DJ, Mayer-Proschel M. Contributions of neurotropic human herpesviruses herpes simplex virus 1 and human herpesvirus 6 to neurodegenerative disease pathology. Neural Regen Res. 2018;13:211–21.
51. 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. 2018;99:64–82.
52. Bourgade K, Garneau H, Giroux G, Le Page AY, Bocti C, Dupuis G, et al. β-Amyloid peptides display protective activity against the human Alzheimer's disease-associated herpes simplex virus-1. Biogerontology. 2014;16:85–98.

Figures

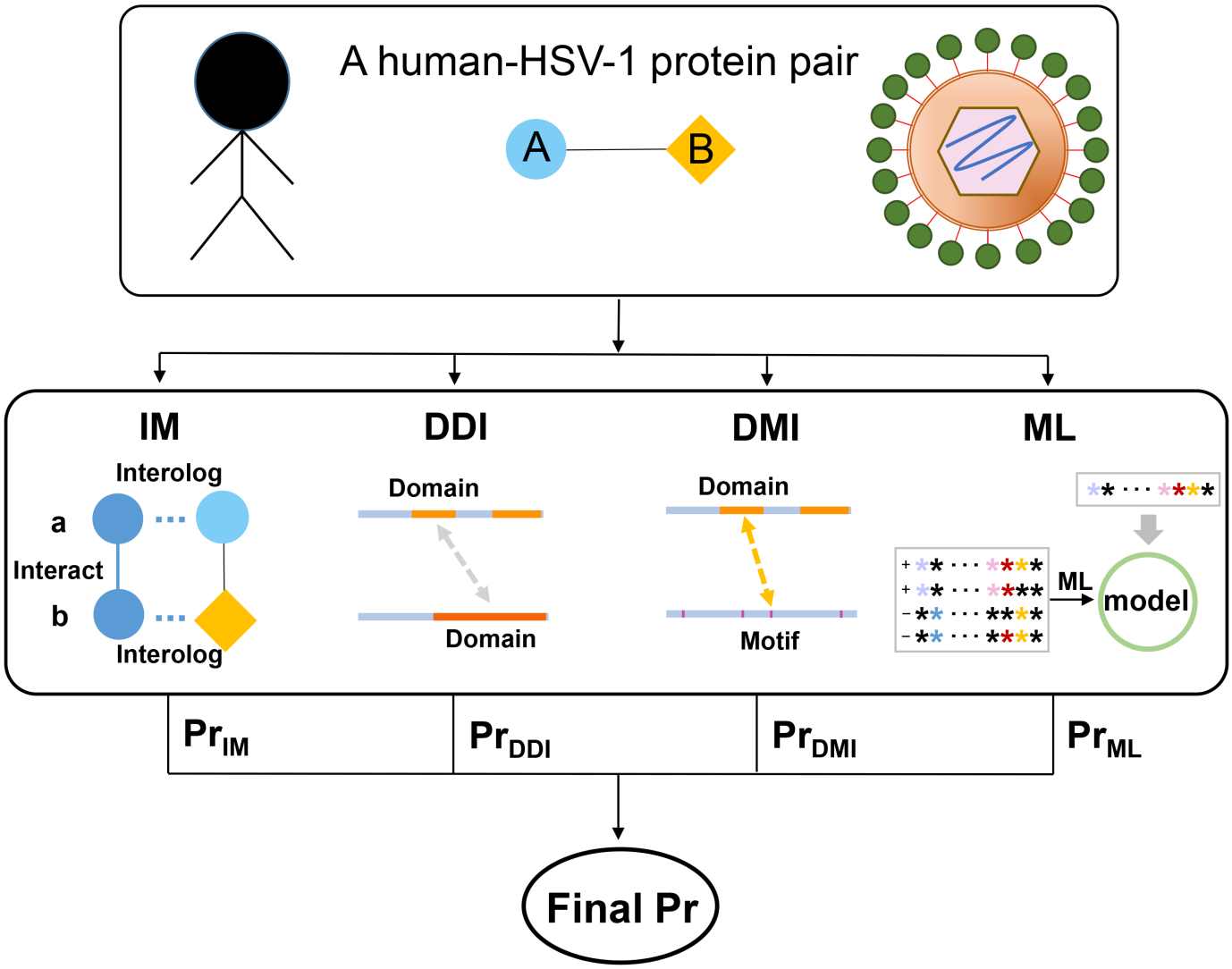

Figure 1

Workflow for prediction of human-HSV-1 PPIs. We first evaluated the interaction probability for each human-HSV-1 protein pair by interolog mapping (IM), domain-domain interaction (DDI), domain-motif interaction (DMI) or machine learning (ML)-based methods. Then, the four interaction probability scores (PrIM, PrDDI, PrDMI and PrML) were integrated into a final probability score (Pr).

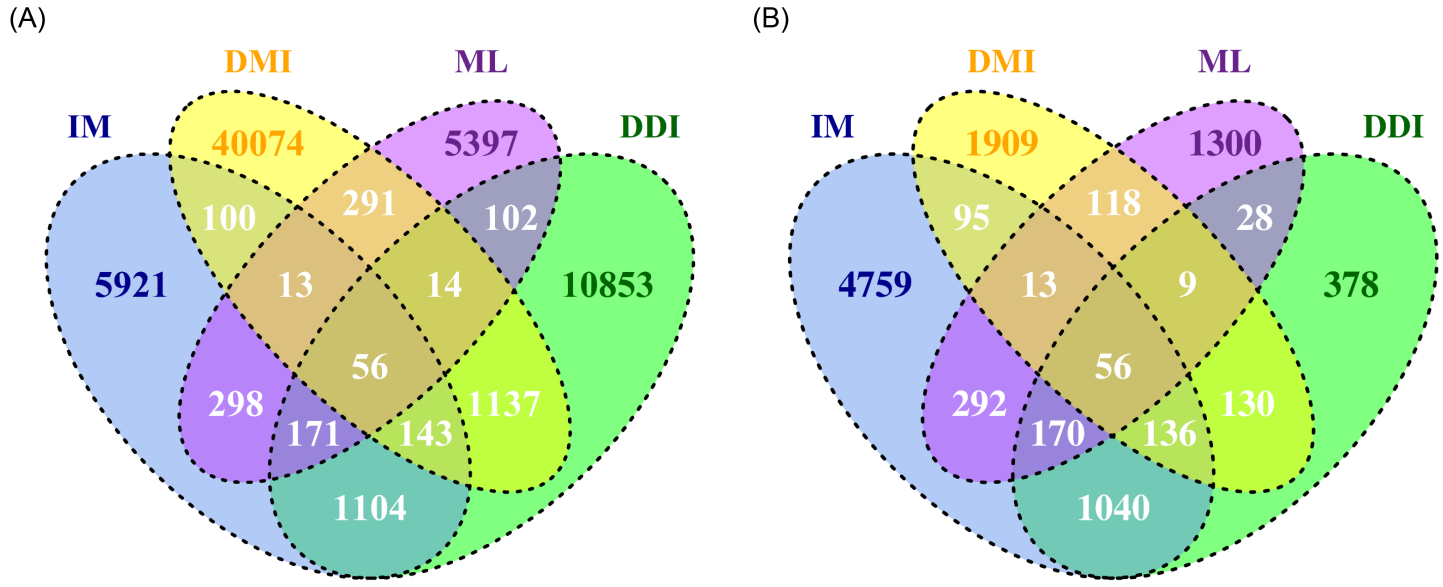

Figure 2

Overlaps of predicted PPIs among four individual methods. (A) All the predictions (Pr > 0); (B) High-confidence predictions (Pr > 0.5).

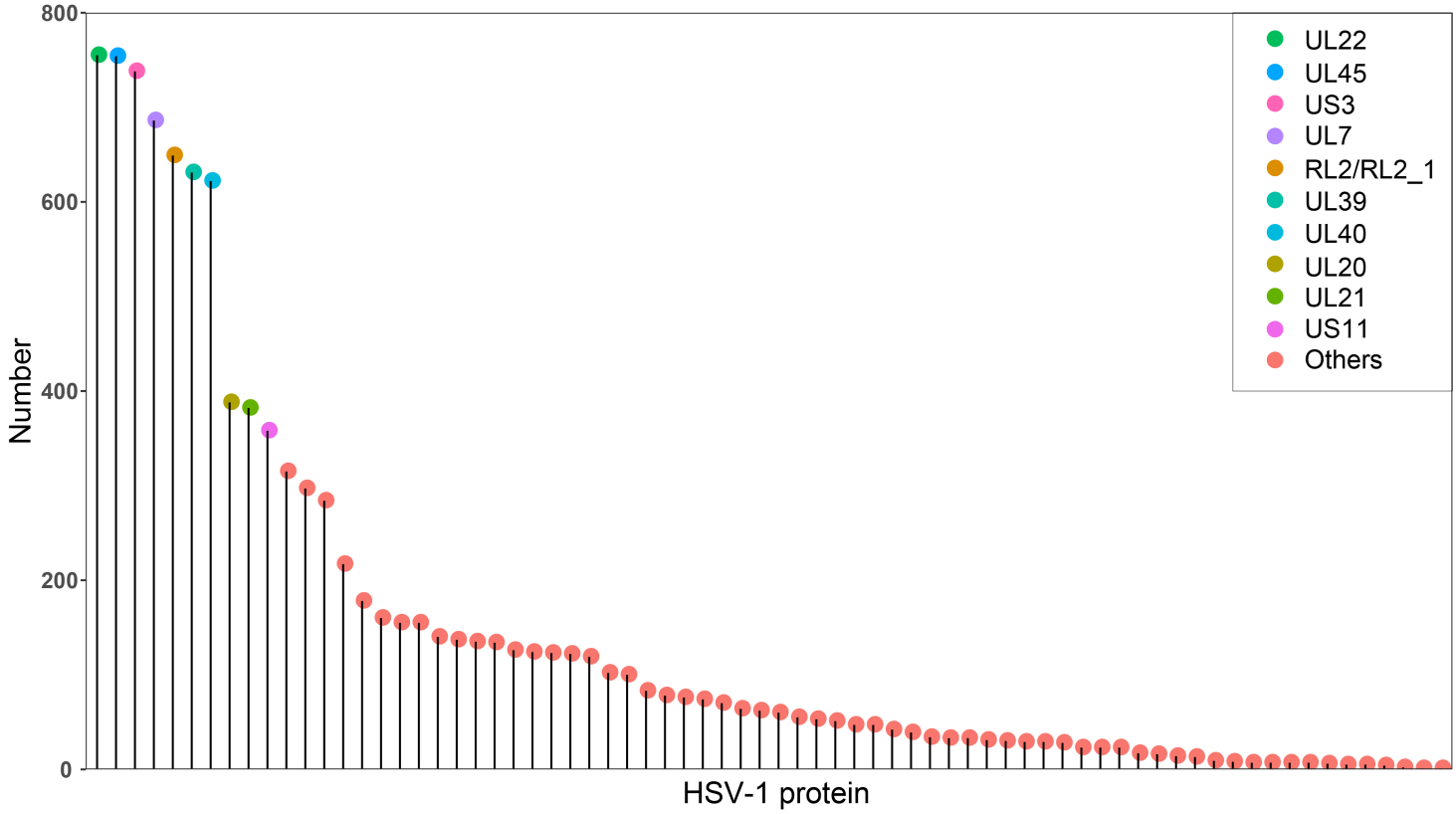

Figure 3

Number of human proteins predicted to interact with HSV-1 proteins.

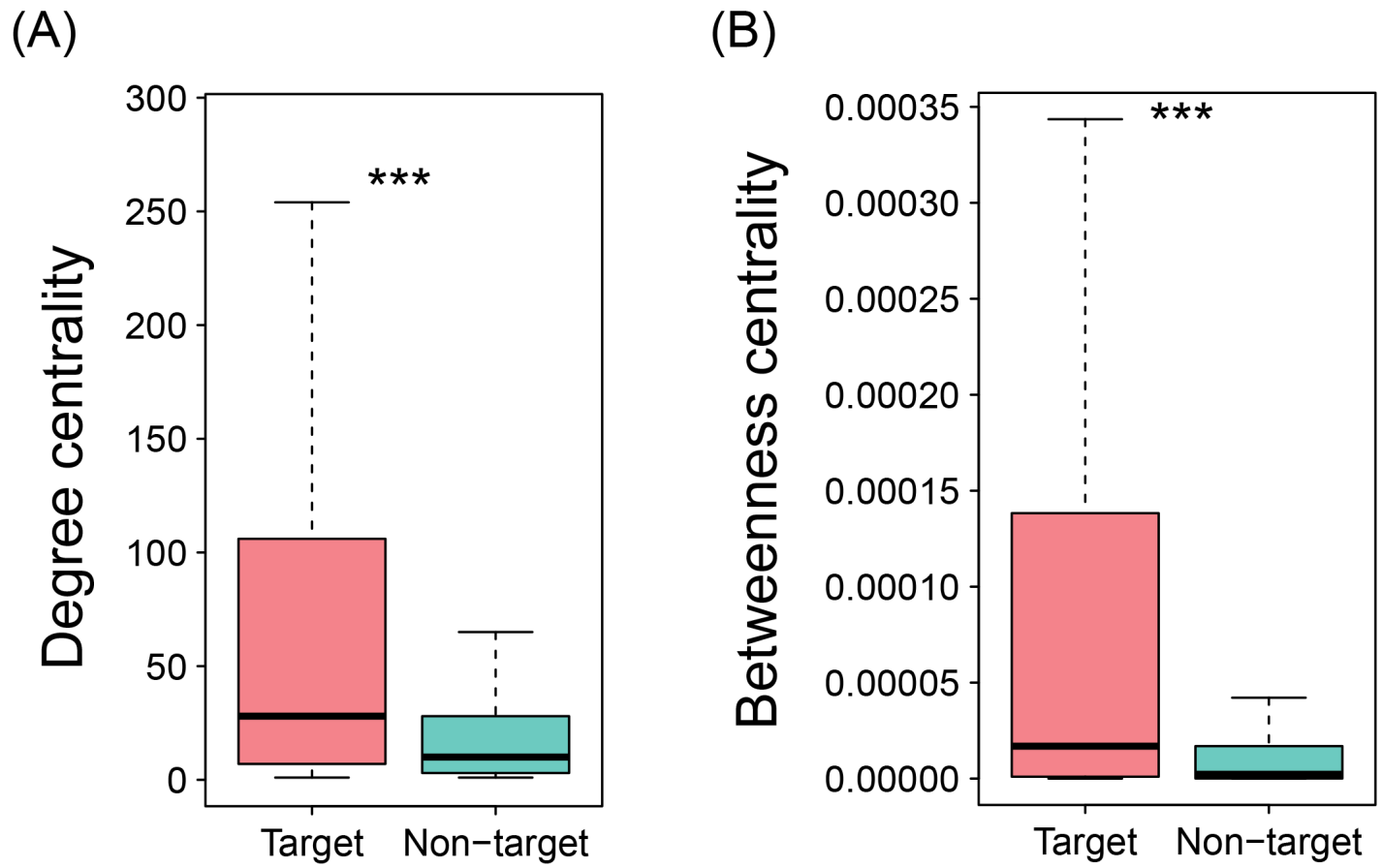

Figure 4

Degree and betweenness centrality of human target proteins and non-target proteins. ***, stands for statistically significant results (Wilcoxon test, P value < 2.2 × 10-16).

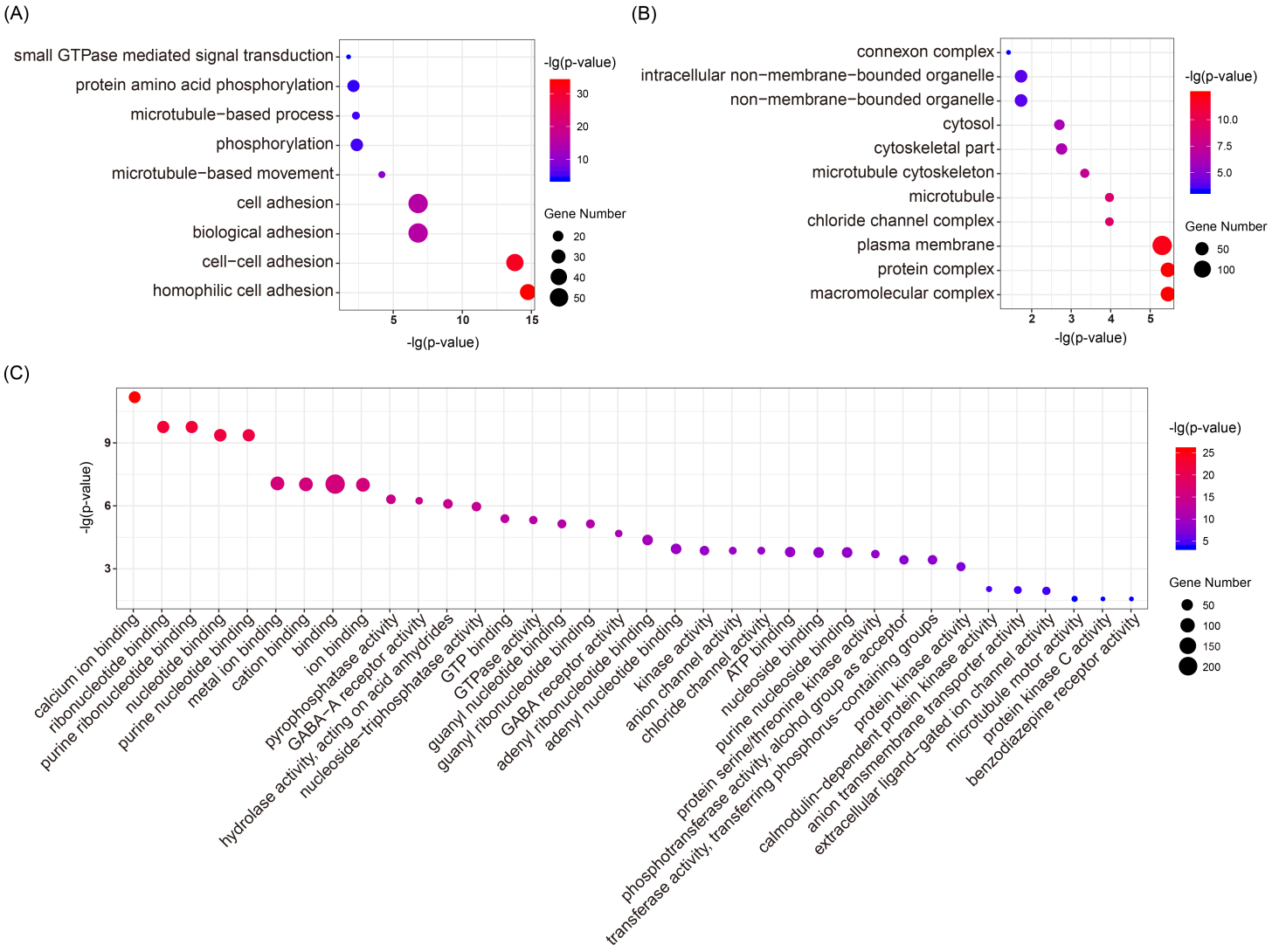

Figure 5

Enriched GO terms of the brain-specific human proteins predicted to interact with HSV-1 proteins in the categories of biological processes (A), cell component (B) and molecular function (C).

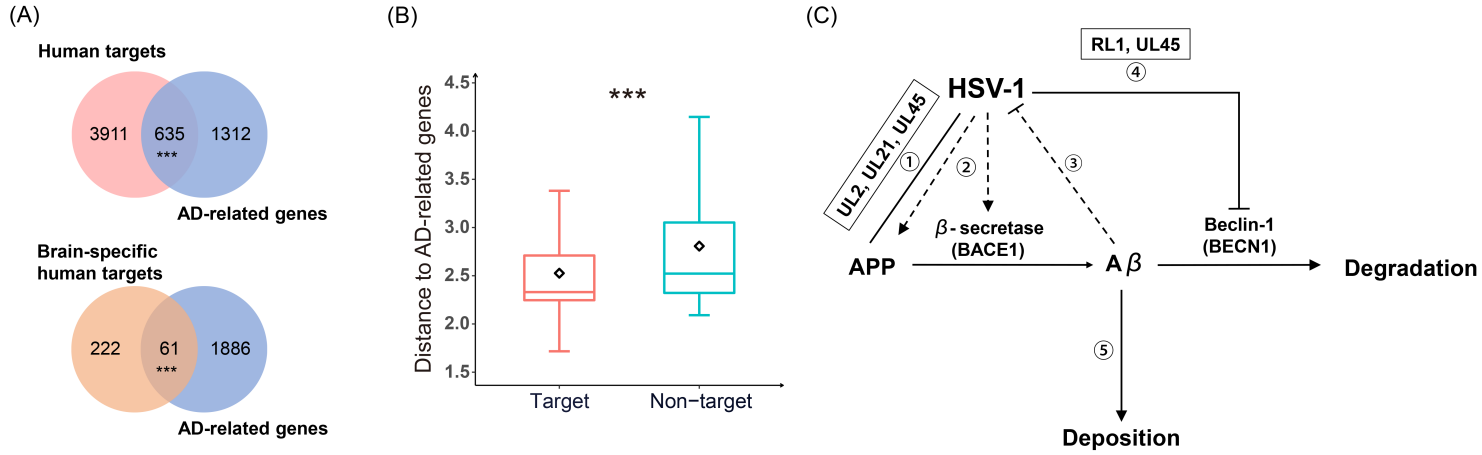

Figure 6

Association between human-HSV-1 PPIs and AD. (A) Overlaps between human targets and AD-related genes. (B) Differences in network distance between target proteins and non-target proteins to AD-related genes in the human PPI network. *** denotes statistically significance (Wilcoxon test, P value < 2.2 × 10-16). The small diamond box represents the mean value. (C) Relationship among APP, Aβ and HSV-1. HSV-1 proteins (UL2, UL21, and UL45) interact with APP. HSV-1 infection induces APP phosphorylation and an increase in the activity of BACE1, resulting in the conversion of APP to Aβ. Aβ inhibits viral activity by encapsulating viral proteins. HSV-1 proteins (RL1 and UL45) interact with Beclin-1 suppressing the degradation of Aβ by inhibiting its autophagy-lysosome pathway. In sum, the imbalance between Aβ production and degradation leads to the deposition of Aβ.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1DatasetS1.xlsx
- Additionalfile3DatasetS3.xlsx
- Additionalfile2DatasetS2.xlsx
- Additionalfile4.pdf