Membrane proteins with high N-glycosylation, high expression, and multiple interaction partners were preferred by mammalian viruses as receptors

Zheng Zhang[1, #], Zhaozhong Zhu[1, #], Wenjun Chen[1], Zena Cai[1], Beibei Xu[2], Zhiying Tan[2], Aiping Wu[3,4], Xingyi Ge[1], Xinhong Guo[1], Zhongyang Tan[1], Zanxian Xia[5], Haizhen Zhu[1, 6, *], Taijiao Jiang[3, 4, *], Yousong Peng[1, *]

[1] College of Biology, Hunan University, Changsha, China

[2] College of Computer Science and Electronic Engineering, Hunan University, Changsha, China

[3] Center of System Medicine, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China

[4] Suzhou Institute of Systems Medicine, Suzhou, China

[5] School of Life Sciences, Central South University, Changsha, China

[6] State Key Laboratory of Chemo/Biosensing and Chemometrics, Hunan University, Changsha, China

# These authors contributed equally to this work

* Correspondence: zhuhaizhen69@yahoo.com (HZ), taijiao@ibms.pumc.edu.cn (TJ), pys2013@hnu.edu.cn (YP)

ABSTRACT

Receptor mediated entry is the first step for viral infection. There are thousands of cell membrane proteins, yet only a few of them were identified as viral receptors. How the virus selects receptors is an unsolved important question. Here, by manually curating a high-quality database of 268 pairs of mammalian virus-host receptor interaction, which included 128 unique viral species or sub-species and 119 virus receptors, we found the viral receptors were structurally and functionally diverse, yet they had several common features when compared to other cell membrane proteins: more protein domains, higher level of N-glycosylation, higher ratio of self-interaction and more interaction partners, and higher expression in most tissues of the host. Additionally, the receptors used by the same virus tended to co-evolve. This work could deepen our understanding towards the viral receptor selection and help for identification of viral receptors.

INTRODUCTION

In the new century, much progress has been made in prevention and control of infectious diseases, but the recent serial outbreaks of Zika virus (Mlakar et al., 2016), Ebola virus (EBOV) (Maganga et al., 2014) and Middle East Respiratory Syndrome Coronavirus (MERS-CoV) (Breban et al., 2013) indicate that the viral infectious diseases still pose a serious threat to human health and global security. The virus is the most abundant biological entity on Earth and exists in all habitats of the world

(Paez-Espino et al., 2016). Nearly all cellular organisms are prey to viral attack. Humans were reported to be infected by hundreds of viruses (Geoghegan et al., 2016; Mihara et al., 2016). Most of the human emerging infectious diseases are zoonotic, with viruses that originate in mammals of particular concern (Olival et al., 2017), such as the Human Immunodeficiency Virus (HIV) (Sharp and Hahn, 2011) and Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) (Ge et al., 2013). Mammals are not only the most closely related animal to humans in phylogeny, but also contact with humans most frequently (Olival et al., 2017), especially for the livestock and pet. For effective control of human viral diseases, much attention should be paid to the mammalian virus.

Receptor-binding is the first step for viral infection of host cells (Baranowski et al., 2001; Dimitrov, 2004; Grove and Marsh, 2011; Li, 2015a). Multiple types of molecules could be used as viral receptors (Baranowski et al., 2001; Casasnovas, 2013), including protein (Li, 2016; Wang, 2002; Yan et al., 2012), carbohydrate (Isa et al., 2006; Peng et al., 2017) and lipid (Mazzon and Mercer, 2014). How to select receptors by the virus is an important unsolved question (Casasnovas, 2013; Grove and Marsh, 2011; Li, 2016; Marija Backovic, 2012). Specificity and affinity are two most important factors for viral receptor selection (Casasnovas, 2013). Carbohydrates and lipids are widely distributed on host cell surfaces and easy targets for viruses to grab (Dimitrov, 2004; Li, 2015a). Compared to these molecules, proteins were reported to be more suitable receptors because of stronger affinity and higher specificity for viral attachment, which could increase the efficiency of viral entry and facilitate viruses to expand their host ranges and alter their tropisms (Baranowski et al., 2001; Casasnovas, 2013; Dimitrov, 2004; Li, 2015a; Wang, 2002). Previous studies have shown that proteins that were abundant in the host cell surface or had relatively low affinity for their natural ligands, were preferred by viruses as receptors, such as proteins involved in cell adhesion and recognition by reversible, multivalent avidity-determined interactions (Dimitrov, 2004; Wang, 2002). This suggests that the selection of proteins by viruses as receptors should not be a random process. A systematic analysis of the characteristics of the viral receptor could help understand the mechanisms under the receptor selection by viruses.

Here, by manually curating a high-quality database of 268 pairs of mammalian virus-receptor interaction, which included 128 unique viral species or sub-species and 119 virus receptors, we systematically analyzed the structural, functional, evolutionary and tissue-specific expression characteristics of mammalian virus receptors, which could not only deepen our understanding towards the mechanism behind the viral receptor selection, but also help to predict and identify viral receptors.

RESULTS

Database of mammalian virus-host receptor interaction

To understand how the virus selects receptors, we manually curated a high-quality database of 268 pairs of mammalian virus-host receptor interactions (Figure 1), which included 128 unique viral species or sub-species from 21 viral families and 119 virus receptors from 13 mammal species. These viruses covered all groups of viruses in the Baltimore classification (Figure 1). Among them, the single-stranded RNA (ssRNA) virus accounted for over half of all viruses (76/128), while the double-stranded RNA (dsRNA) virus accounted for the least (3/128). On the level of family, the family of *Picornaviridae* of ssRNA virus, *Retroviridae* of Retro-transcribing viruses (RT) and *Herpesviridae* of double-stranded DNA (dsDNA) viruses were the most abundant ones in the database (Figure 1 and Table S1).

The viral receptor collected here belonged to 13 mammal species (Figure S1A), among which the human accounted for the most (74/119). Analysis of the association between the virus and their receptors showed that 60% of viruses (77/128) used only one receptor (Figure 1 and Figure S1B). Surprisingly, some viruses, such as the Human alphaherpesvirus 1 (HSV-1) and Hepacivirus C (HCV), used more than five receptors. We next analyzed the receptor usage on the level of viral family and found that different viruses of the same family usually used different receptors. For example, in the family of *Togaviridae*, the Chikungunya virus (CHIKV), the Rubella virus (RBV) and the Sindbis virus (SBV) used the receptor of prohibitin (PHB), myelin oligodendrocyte glycoprotein (MOG) and ribosomal protein SA (RPSA), respectively. On the other hand, some viruses of different families or even different groups used the same receptor (Figure 1). For example, HIV-2 and EBOV, from the family of *Retroviridae* (RT group) and *Filoviridae* (ssRNA group) respectively, both took CD209 molecule (CD209) (marked with an asterisk in Figure 1) as the receptor. On average, each receptor was used by more than two viruses. More specifically, among

119 virus receptors, forty-four of them were used by more than one virus (Figure 1 and Figure S1C); twenty-one of them were used by viruses of more than one family and fifteen of them were used by viruses of more than one group (Figure 1).

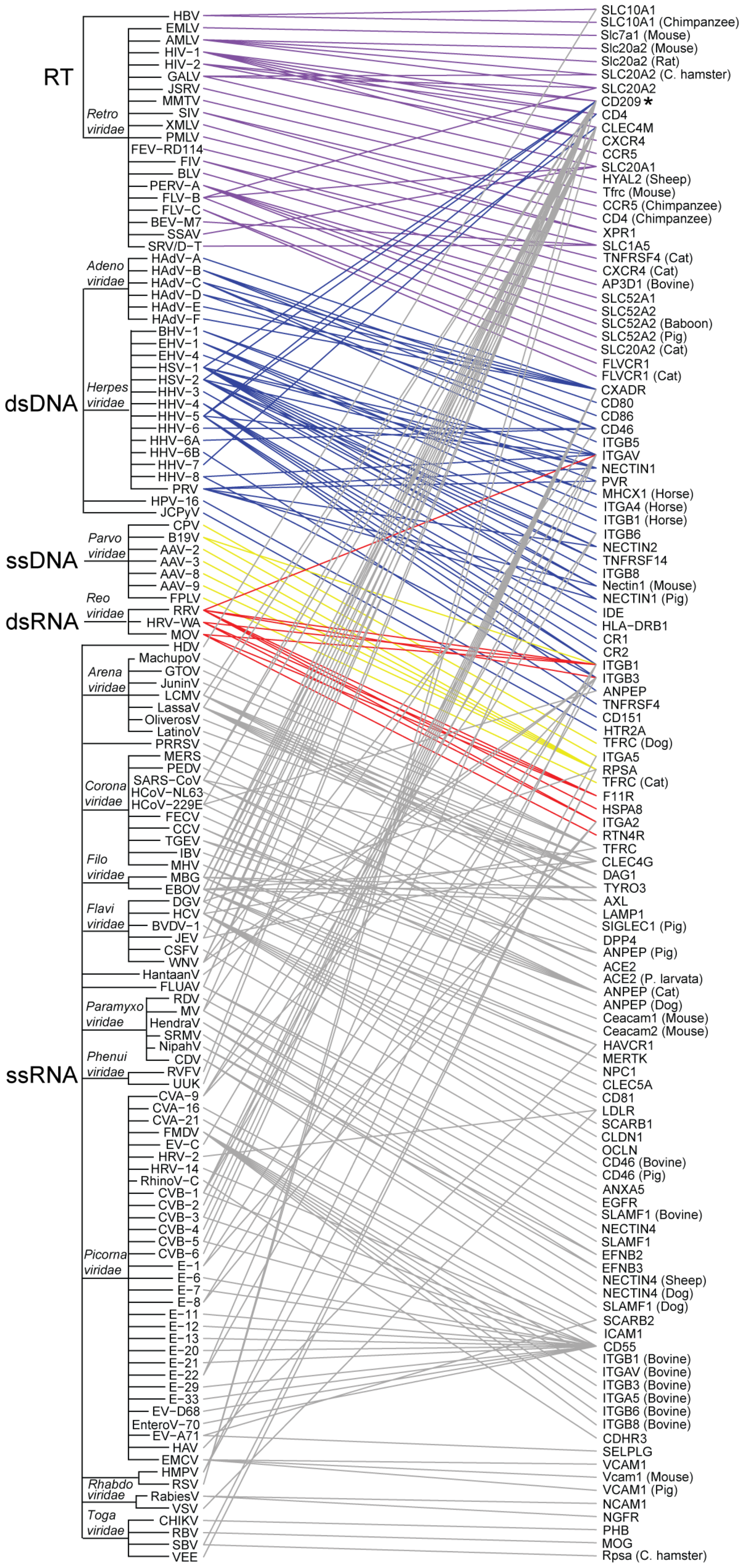

RT
Retro viridae
HBV
EMLV
AMLV
HIV−1
HIV−2
GALV
JSRV
MMTV
SIV
XMLV
PMLV
FEV−RD114
FIV
BLV
PERV−A
FLV−B
FLV−C
BEV−M7
SSAV
SRV/D−T
dsDNA
Adeno viridae
HAdV−A
HAdV−B
HAdV−C
HAdV−D
HAdV−E
HAdV−F
Herpes viridae
BHV−1
EHV−1
EHV−4
HSV−1
HSV−2
HHV−3
HHV−4
HHV−5
HHV−6
HHV−6A
HHV−6B
HHV−7
HHV−8
PRV
HPV−16
JCPyV
ssDNA
Parvo viridae
CPV
B19V
AAV−2
AAV−3
AAV−8
AAV−9
FPLV
dsRNA
Reo viridae
RRV
HRV−WA
MOV
ssRNA
HDV
Arena viridae
MachupoV
GTOV
JuninV
LCMV
LassaV
OliverosV
LatinoV
PRRSV
Corona viridae
MERS
PEDV
SARS−CoV
HCoV−NL63
HCoV−229E
FECV
CCV
TGEV
IBV
MHV
Filo viridae
MBG
EBOV
Flavi viridae
DGV
HCV
BVDV−1
JEV
CSFV
WNV
HantaanV
FLUAV
Paramyxo viridae
RDV
MV
HendraV
SRMV
NipahV
CDV
Phenui viridae
RVFV
UUK
Picorna viridae
CVA−9
CVA−16
CVA−21
FMDV
EV−C
HRV−2
HRV−14
RhinoV−C
CVB−1
CVB−2
CVB−3
CVB−4
CVB−5
CVB−6
E−1
E−6
E−7
E−8
E−11
E−12
E−13
E−20
E−21
E−22
E−29
E−33
EV−D68
EnteroV−70
EV−A71
HAV
EMCV
HMPV
RSV
Rhabdo viridae
RabiesV
VSV
Toga viridae
CHIKV
RBV
SBV
VEE
SLC10A1
SLC10A1 (Chimpanzee)
Slc7a1 (Mouse)
Slc20a2 (Mouse)
Slc20a2 (Rat)
SLC20A2 (C. hamster)
SLC20A2
CD209 *
CD4
CLEC4M
CXCR4
CCR5
SLC20A1
HYAL2 (Sheep)
Tfrc (Mouse)
CCR5 (Chimpanzee)
CD4 (Chimpanzee)
XPR1
SLC1A5
TNFRSF4 (Cat)
CXCR4 (Cat)
AP3D1 (Bovine)
SLC52A1
SLC52A2
SLC52A2 (Baboon)
SLC52A2 (Pig)
SLC20A2 (Cat)
FLVCR1
FLVCR1 (Cat)
CXADR
CD80
CD86
CD46
ITGB5
ITGAV
NECTIN1
PVR
MHCX1 (Horse)
ITGA4 (Horse)
ITGB1 (Horse)
ITGB6
NECTIN2
TNFRSF14
ITGB8
Nectin1 (Mouse)
NECTIN1 (Pig)
IDE
HLA−DRB1
CR1
CR2
ITGB1
ITGB3
ANPEP
TNFRSF4
CD151
HTR2A
TFRC (Dog)
ITGA5
RPSA
TFRC (Cat)
F11R
HSPA8
ITGA2
RTN4R
TFRC
CLEC4G
DAG1
TYRO3
AXL
LAMP1
SIGLEC1 (Pig)
DPP4
ANPEP (Pig)
ACE2
ACE2 (P. larvata)
ANPEP (Cat)
ANPEP (Dog)
Ceacam1 (Mouse)
Ceacam2 (Mouse)
HAVCR1
MERTK
NPC1
CLEC5A
CD81
LDLR
SCARB1
CLDN1
OCLN
CD46 (Bovine)
CD46 (Pig)
ANXA5
EGFR
SLAMF1 (Bovine)
NECTIN4
SLAMF1
EFNB2
EFNB3
NECTIN4 (Sheep)
NECTIN4 (Dog)
SLAMF1 (Dog)
SCARB2
ICAM1
CD55
ITGB1 (Bovine)
ITGAV (Bovine)
ITGB3 (Bovine)
ITGA5 (Bovine)
ITGB6 (Bovine)
ITGB8 (Bovine)
CDHR3
SELPLG
VCAM1
Vcam1 (Mouse)
VCAM1 (Pig)
NCAM1
NGFR
PHB
MOG
Rpsa (C. hamster)

Figure 1. The mammalian viruses and their related receptors in our database. The lines between the virus and their related receptors were colored according to the group of the virus in the Baltimore classification. The names of some viral families were presented in italic. Viral names were displayed in abbreviation (see Table S1 for the full name). The host names were given for the receptor of non-human mammal species. The receptor CD209 was marked with an asterisk. See also Table S1 and Figure S1.

Structural, functional, evolutionary and tissue-specific expression characteristics of mammalian virus receptors

To understand how the virus selects receptors, we systematically analyzed the structural, functional, evolutionary and tissue-specific expression characteristics of the mammalian virus receptor.

1) The mammalian virus receptor were structurally diverse

We firstly investigated the structural characteristics of mammalian virus receptor proteins. As expected, all the mammalian virus receptor protein belonged to the membrane protein which had at least one transmembrane alpha helix (Figure S2A). Twenty-four of them had more than five helixes, such as 5-hydroxytryptamine receptor 2A (HTR2A) and NPC intracellular cholesterol transporter 1 (NPC1). The receptor protein was mainly located in the cell membrane. Besides, more than one third (43/119) of them were also located in the cytoplasm, and thirteen of them were located in the nucleus.

Then, the protein domain composition of the mammalian virus receptor protein was analyzed. The mammalian virus receptor proteins contained a total of 336 domains based on the Pfam database, with each viral receptor protein containing more than two domains on average (Figure S2B). This was significantly more than that of human proteins or human membrane proteins (p-values < 0.001 in the Wilcoxon rank-sum test). Some viral receptor proteins may contain more than 10 domains, such as complement C3d receptor 2 (CR2) and low density lipoprotein receptor (LDLR). The protein domains of the mammalian virus receptor protein could be grouped into 77 families in the Pfam database, suggesting the structure diversity of the mammalian virus receptor protein. The most commonly observed Pfam families were Immunoglobulin V-set domain, Immunoglobulin C2-set domain, Integrin beta chain VWA domain, Integrin plexin domain, and so on (Figure S2C).

2) The mammalian virus receptor had high level of N-glycosylation

Glycosylation of protein is widespread in the eukaryote cell. We next characterized the glycosylation level of the mammalian virus receptor. N-glycosylation is the most common type of glycosylation. We found that 93 of 119 mammalian virus receptors were N-glycosylated with an average of 0.94 glycosylation sites per 100 amino acids (Figure 2A). It increased to 0.97 glycosylation sites per 100 amino acids for the human viral receptor (Figure 2A), among which 62 were N-glycosylated. Twelve human viral receptors were observed to have ten or more N-glycosylation sites, such as complement C3b/C4b receptor 1 (CR1) and lysosomal associated membrane protein 1 (LAMP1). Figure 2B displayed the modeled 3D-structure of HTR2A, the receptor for JC polyomavirus (JCPyV). Five N-glycosylation sites were highlighted in red on the structure, which were reported to be important for viral infection (Maginnis et al., 2010). For comparison, we also characterized the N-glycosylation level for the human cell membrane protein, human membrane proteins and all human proteins (Figure 2A). It was found they had a significantly lower level of N-glycosylation than that of human and mammalian virus receptors (p-values < 0.001 in the Wilcoxon rank-sum test), which suggests the importance of N-glycosylation for the viral receptor.

O-glycosylation is also a common type of glycosylation. We found there was only a small fraction of mammalian virus receptors (14/119) with O-glycosylation. Besides, no significant difference was observed between the O-glycosylation level of mammalian virus receptor proteins and that of human proteins (Figure S2D).

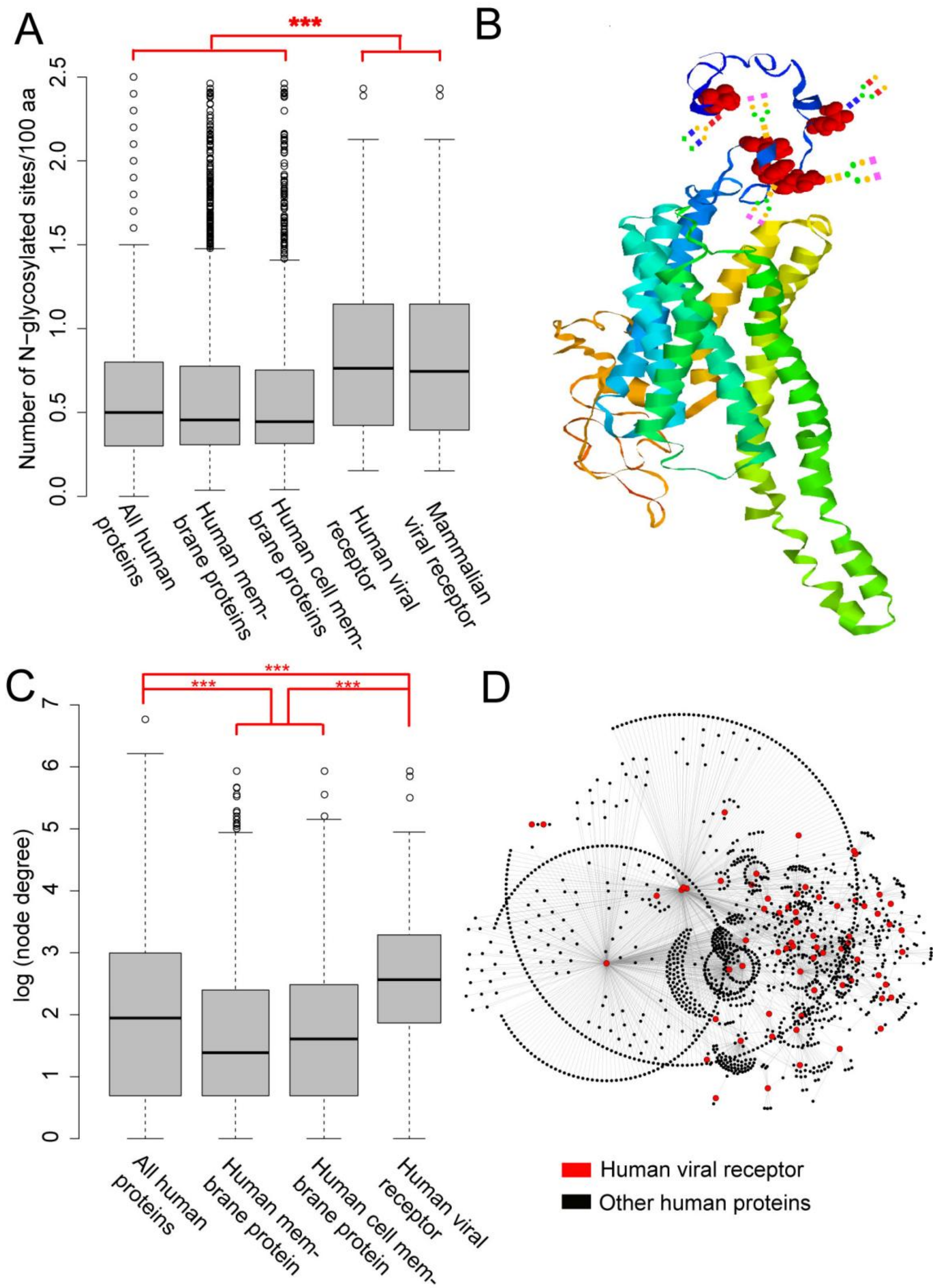


Figure 2. Analysis of N-glycosylation and protein-protein interactions of mammalian virus receptors. (A) Comparison of the N-glycosylation level between mammalian viral receptors, human viral receptors, human cell membrane proteins, human

membrane proteins and all human proteins. For clarity, the outliers greater than 2.5 were removed. “***”, p-value < 0.001. (B) The modeled 3D-structure of HTR2A. Five N-glycosylation sites were highlighted in red. Artificial glycans were manually added onto the site. (C) Comparison of the degree of proteins between human viral receptors, human cell membrane proteins, human membrane proteins and all human proteins in the human PPIN. For clarity, the node degree was logarithmically transformed. “***”, p-value < 0.001. (D) Partial human PPI network composing of the PPIs which involved at least one viral receptor protein (colored in red). See also Table S2&S3 and Figure S2&S3.

3) Functional enrichment analysis of the human virus receptor

We next attempted to identify the gene functions and pathways enriched in the mammalian virus receptor. As was mentioned above, 74 of 119 mammalian virus receptors belonged to the human. Besides, analysis showed that 36 of the remaining non-human mammalian virus receptors were homologs of the human virus receptor (Table S2). Therefore, we conducted the function enrichment analysis only for the human virus receptor based on the databases of Gene Ontology (GO) and KEGG. For the GO Cellular Component (Table S3), the human virus receptor was mainly enriched in the membranes and junctions, the latter of which included the adherens junction, cell-substrate junction, focal adhesion, and so on. For the GO Biological Process (Table S3), the human virus receptor was mainly enriched in the process of entry into the host. Besides, some terms related to the immune response were also enriched, such as “Regulation of leukocyte activation” and “Lymphocyte activation”. For the GO Molecular Function (Table S3), besides for the enrichment of terms related to the virus receptor activity, the human virus receptor was also enriched in terms of binding to integrin, glycoprotein, cytokine, and so on.

Consistent with the enrichment analysis of GO Cellular Component, the KEGG pathways of “Cell adhesion molecules”, “Focal adhesion” and “ECM-receptor interaction” were also enriched. Besides, the pathway of “Phagosome” was enriched (Table S3), which may be associated with viral entry into the host cell. Interestingly, some pathways associated with heart diseases were enriched, including “Dilated cardiomyopathy”, “Hypertrophic cardiomyopathy”, “Arrhythmogenic right ventricular cardiomyopathy” and “Viral myocarditis”.

4) Human virus receptors had more interaction partners than other proteins

We next analyzed the protein-protein interactions (PPIs) which the mammalian virus receptor protein took part in. As the reason mentioned above, we only used the human virus receptor for PPI analysis. A human PPI network (PPIN) was constructed based on the work of Menche et al (Menche et al., 2015). It included a total of 13,460 human proteins that are interconnected by 141,296 interactions. The degree and betweenness of each protein in the PPIN were calculated, which could measure the importance of a protein in the PPIN. It was found that the degrees for human membrane proteins and cell membrane proteins were significantly smaller than those of other human proteins (Figure 2C & Figure S3A, p-value < 0.001 in Wilcox rank-sum test) in the PPIN. Similar observations could be found for the node betweenness in the PPIN (Figure S3B&C). However, the human virus receptor protein, a subset of the human cell membrane protein, was found to have significantly larger degrees and higher betweenness than other human proteins in the PPIN (Figure 2C and Figure S3, p-value < 0.001 in Wilcox rank-sum test). The median degrees for human virus receptors was 13 (Figure 2C), which was nearly twice as much as that of all human proteins. Six viral receptors were observed to have degrees larger than 100 (Figure 2D), including epidermal growth factor receptor (EGFR), heat shock protein family A member 8 (HSPA8), PHB, RPSA, CD4 molecule (CD4) and integrin subunit beta 1 (ITGB1). Since the viral receptor (colored in red in Figure 2D) interacted with lots of other human proteins (colored in black in Figure 2D) in PPIN, we further investigated the functional enrichment of these proteins by GO enrichment analysis. Interestingly, six of top ten enriched terms in the domain of Biological Process were related to protein targeting or localization (Table S3).

When looking at the interactions between viral receptors, we found that 38 of 74 viral receptors interacted with themselves. This ratio (38/74 = 51%) was much higher than that of human proteins (22%), membrane proteins (11%) and human cell membrane proteins (14%). However, we found the viral receptor tended not to interact with each other (Figure S3D). Among 74 human virus receptor proteins, 36 of them had no interactions with any other human virus receptor. There were only 50 PPIs between different human virus receptor proteins, with each viral receptor protein interacting with an average of only one other viral receptor protein.

5) The mammalian virus receptor was not more conserved than other genes

Large degree of the human viral receptor in the human PPIN suggests the importance of them in cellular activity. Analysis showed that 11 human viral receptors belonged to the housekeeping gene. This ratio (0.15 = 11/74) was a little lower than that of housekeeping genes in all human genes (0.19 = 3804/20243), suggesting that the human viral receptor was not enriched in the housekeeping gene.

Then, we investigated the evolutionary conservation of mammalian virus receptors in 108 mammal species which were richly annotated in the NCBI Reference Sequences (RefSeq) database (see EXPERIMENTAL PROCEDURES). Over half of mammalian virus receptors had homologs (see EXPERIMENTAL PROCEDURES) in more than 100 mammal species (Figure S4A). We further calculated the pairwise sequence identities between the viral receptor and their homologs in mammal species. For nearly half of viral receptors, the average of pairwise sequence identities was higher than 0.8 (Figure S4B). For comparison, we also analyzed the evolutionary conservation of human proteins by randomly selecting 1000 human proteins from the NCBI RefSeq database. They were observed to have similar conservation level with that of human viral receptors (Figure S4C&D and Table S4).

6) Viral receptors expressed higher than other proteins in 32 major human tissues

Since the virus has to compete with other proteins for binding to the receptor, proteins with high expression should be preferred by viruses as receptors. Thus we measured the average expression level of human viral receptors in 32 major human tissues (Figure 3). For comparison, those of human membrane genes, human cell membrane genes and all human genes were also displayed in Figure 3. As was shown clearly, the expression level of human cell membrane genes (in cyan) was similar to that of all human genes (in black) in these tissues. Both of them were lower than that of human membrane genes (in blue). However, the human viral receptor (in red), which was part of the human cell membrane gene, expressed much higher than other sets of genes in nearly all these tissues. On average, they had an expression level of 24 transcripts per million (TPM) in these tissues, while this was 8, 4 and 4 TPM for the human membrane gene, the human cell membrane gene and all human genes (see the black arrow in Figure 3).

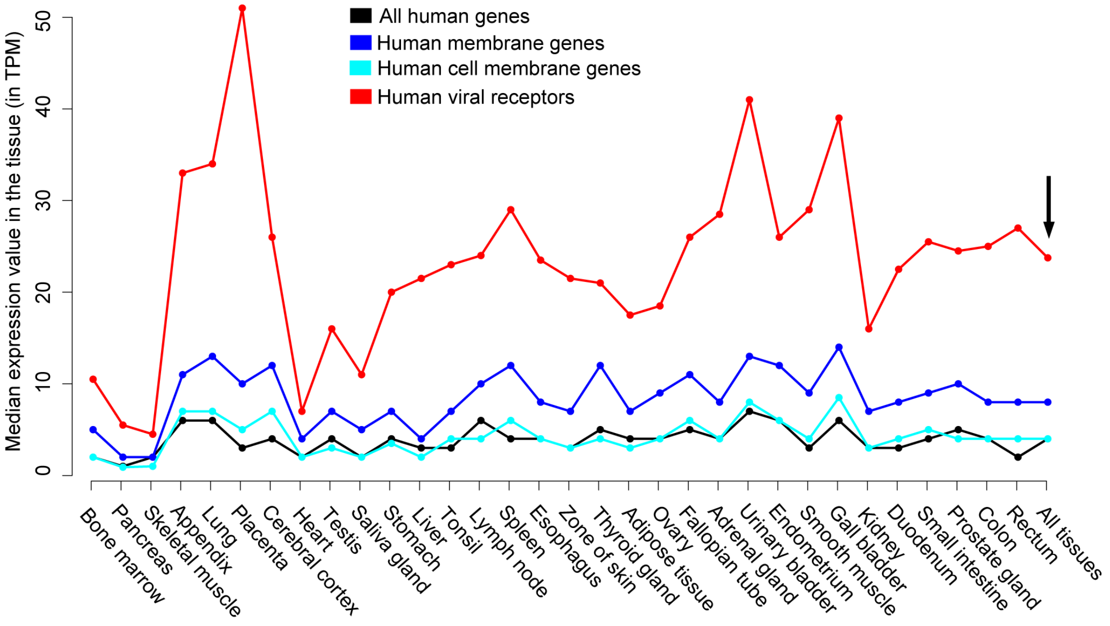


Figure 3. The average expression level of human viral receptors (red), human cell membrane genes (cyan), human membrane genes (blue) and all human genes (black) in 32 major human tissues. The expression level was measured with transcripts per

million (TPM). The black arrow refers to the average expression level of genes in all 32 tissues.

Viral receptors used by the same virus tended to co-evolve

The results mentioned above shows that a total of 51 viruses used more than one viral receptor. These viral receptors may work together or independently. We then analyzed the relationship between them. Structural analysis shows that except for integrins which generally work in heterodimer, few of viral receptors used by the same virus shared the same protein domain (data not shown). This suggests that when the virus expands the use of receptors, it tends to select structurally diverse proteins. We continued to analyze the co-evolution between mammalian viral receptors in 108 mammal species. We found that the average of Spearman Correlation Coefficients (SCCs, a measure of the extent of co-evolution) between viral receptors employed by the same virus was 0.54, which was significantly larger than that between other viral receptors (Figure 4A, p-value < 0.001 in the Wilcoxon rank-sum test). For example, SARS-CoV used four receptors, i.e., ACE2, CD209, C-type lectin domain family 4 member G (CLEC4G) and member M (CLEC4M). The average of SCCs between these four receptors was as high as 0.86. In addition, we analyzed the extent of co-expression between human viral receptors in 32 tissues. It was found that the extent of co-expression between viral receptors employed by the same virus was a little larger than that between other viral receptors, yet this difference was not statistically significant (Figure 4B, p-value > 0.1 in the Wilcoxon rank-sum test).

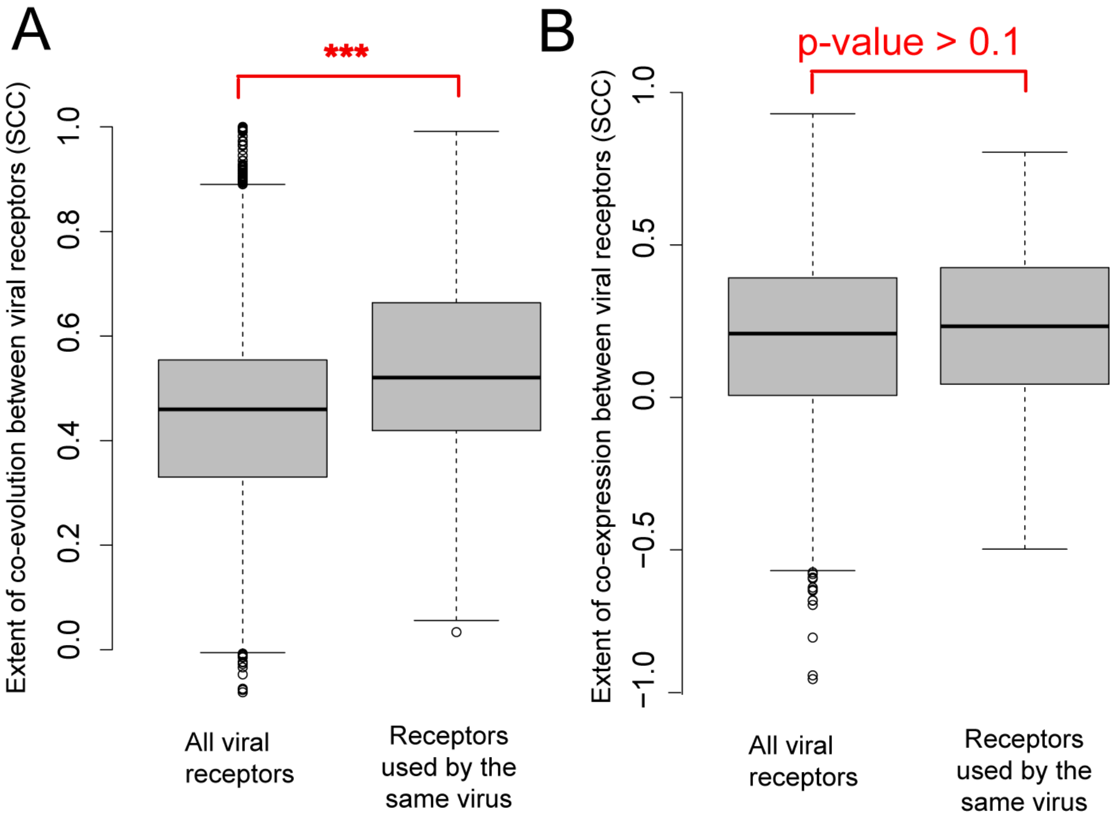


Figure 4. The co-evolution and co-expression of viral receptors. (A) Comparing the extent of co-evolution between mammalian virus receptors in 108 mammal species in the set of receptors used by the same virus and all mammalian virus receptors. “***”, p-value < 0.001 in the Wilcoxon rank-sum test. (B) Comparing the extent of co-expression between human viral receptors in 32 human tissues in the set of receptors used by the same virus and all human virus receptors. See also Table S4&S5 and Figure S4.

DISCUSSION

The viral receptor is essential for viral infection. By collecting the largest dataset ever reported about the mammalian virus-host receptor interactions, we systematically analyzed the structural, functional, evolutionary and tissue-specific expression characteristics of mammalian virus receptors. We found that the viral receptors were a subset of structurally and functionally diverse cell membrane proteins. They were enriched in GO terms and KEGG pathways related to junctions, adhesion and binding, which were typical features of viral receptors reported by previous studies (Baranowski et al., 2001; Casasnovas, 2013; Dimitrov, 2004; Marija Backovic, 2012; Wang, 2002). Besides, our analysis identified several novel features of the viral receptor. Firstly, the viral receptor had a higher level of N-glycosylation than other proteins. Then, what's the relationship between glycosylation and viral receptor selection? As we know, glycosylation of proteins is widely observed in eukaryote cells (Corfield, 2017). It plays an important role in multiple cellular activities, such as folding and stability of glycoprotein, immune response, cell-cell adhesion, and so on. Glycans are abundant on host cell surfaces. They were probably the primordial and fallback receptors for the virus (Li, 2015a). To use glycans as their receptors, a large number of viruses have stolen a host galectin and employed it as a viral lectin (Krupovic and Koonin, 2017; Li, 2015a). For example, the SJR fold, which was mainly responsible for glycan recognition and binding in cellular proteins, was observed in viral capsid proteins of over one fourth of viruses (Krupovic and Koonin, 2017). Thus, during the process of searching for protein receptors, the protein with high level of glycosylation could provide a basal attachment ability for the virus, and should be the preferred receptor for the virus.

Secondly, our analysis showed that the viral receptor protein had a tendency to interact with itself and had far more interaction partners than other membrane proteins. Besides the function of viral receptor, the receptor protein functions in the host cell by interacting with other proteins of the host, such as signal molecules and ligands. Therefore, the virus has to compete with these proteins for binding to the receptor (Wang, 2002). The protein with less interaction partners are expected to be preferred by the virus. Why did the virus select the proteins with multiple interaction partners as receptors? One possible reason is that the receptor proteins are closely related to the "door" of the cell, so that many proteins have to interact with them for in-and-out of the cell. This could be partly validated by the observation that for the interaction partners of human viral receptors, six of top ten enriched terms in the domain of GO Biological Process were related to protein targeting or localization (Table S3). For entry into the cell, the virus also selects these proteins as receptors. Another possible reason is that viral entry into the cell needs cooperation of multiple proteins which were not identified as viral receptors yet. Besides, previous studies show that the virus could structurally mimic native host ligands (Drayman et al., 2013), which help them bind to the host receptor. Thus, membrane proteins with multiple interaction partners have a larger probability to be used by viruses as receptors than other proteins.

Thirdly, the viral receptor was observed to have a much higher level of expression than other genes in each of the 32 human tissues. This may be directly related to the above finding that the viral receptor generally had multiple interaction partners: on the one hand, the viral receptor needs multiple copies to interact with multiple proteins;

on the other hand, since the virus has to compete with other proteins for binding to the receptor, high expression of the receptor will facilitate the virus's binding to the viral receptor.

There were some limitations within the study. Firstly, the viral receptor was biased towards the human, due to the bias of studies towards human viruses. Fortunately, the viral receptor was conserved in mammal species to a large extent, which may reduce the influence of this bias on the diversity of viral receptors. Secondly, due to the difficulties of identifying viral receptors (Li, 2015b; Pillay et al., 2016; Yan et al., 2012), the database of mammalian virus-host receptor interaction was still limited in its size, which hindered us from a more comprehensive survey of the correlation characteristics between viruses and viral receptors. More effective methods, either experimental or computational (Drayman et al., 2013), should be developed for identifying viral receptors, while the characteristics identified in this study may help such endeavors.

Overall, the structural, functional, evolutionary and tissue-specific expression characteristics identified here should not only deepen our understanding of the viral receptor selection, but also help for development of more effective methods for identifying viral receptors.

EXPERIMENTAL PROCEDURES

Database of mammalian virus-host receptor interaction

The data of mammalian virus-host receptor interaction were compiled from three sources: firstly, the literature related to viral receptors (a total of 1303 papers) were downloaded from NCBI Pubmed database (Agarwala et al., 2016) by searching “virus receptor" [TIAB] or "viral receptor"[TIAB] on August 14th, 2017. The mammalian virus and their related host receptors were manually extracted from the literature; secondly, part of viral receptors were directly obtained from the database of ViralZone (Masson et al., 2013) on September 9th, 2017; thirdly, proteins annotated with one of GO terms “virus receptor activity”, “viral entry into host cell” and “receptor activity” in UniprotKB database (Bateman et al., 2017) were collected on August 14th, 2017, and manually checked later. In combination, a database was created with 268 pairs of mammalian virus-host receptor interaction, which included 128 unique viral species or sub-species and 119 viral receptors (Table S1).

Analysis of structural features of viral receptors

The number of transmembrane alpha helix of the mammalian virus receptor was derived from the database of UniprotKB and the web server TMpred (Hofmann and Stoffel, 2017). The location for the viral receptor was inferred from the description of “Subcellular location” for the receptor protein provided by UniProtKB, or from the GO annotations for them: the viral receptors annotated with GO terms which included the words of “cell surface” or “plasma membrane” were considered to be located in the cell membrane; those annotated with GO terms which included the words of “cytoplasm”, “cytosol” or “cytoplasmic vesicle”, or shown to be in the cytoplasm in UniProtKB, were considered to be located in the cytoplasm; those annotated with GO terms "nucleus" (GO:0005634) or "nucleoplasm" (GO:0005654) were considered to be located in the nuclear. The Pfam family, the N-glycosylation and O-glycosylation sites for the protein were obtained from the database of UniprotKB.

For comparison, the human proteins and their related structural characteristics were obtained from the database of UniProtKB/SwissProt on November 24th, 2017. The proteins which had at least one transmembrane alpha helix were considered as membrane proteins. The membrane proteins which were shown to be located in the cell membrane were considered as cell membrane proteins. In total, we obtained 20243 human proteins, 5187 human membrane proteins and 2208 human cell membrane proteins.

The 3D structure for the viral receptor HTR2A were modeled with the help of I-TASSER (Roy et al., 2010) based on the protein sequence of HTR2A (accession number in the database of UniProt: P28223). The best model was selected, and visualized in RasMol (version 2.7.5) (Bernstein, 2017).

Functional enrichment analysis

The GO function and KEGG pathway enrichment analysis for the human viral receptor were conducted with functions of *enrichGO()* and *enrichKEGG()* in the package "clusterProfiler" (version 3.4.4) (Yu et al., 2012) in R (version 3.4.2) (Team, 2016).

Protein-protein interaction (PPI) network analysis

The human PPI network (PPIN) was constructed based on the work of Menche et al

(Menche et al., 2015). The degree and betweenness for each protein in the PPIN were calculated with functions of *degree()* and *betweenness()* in the package "igraph" (version 1.0.0) (G and T, 2006) in R (version 3.2.5). The network was displayed with Cytoscape (version 2.6.2) (Shannon et al., 2003).

For robustness of the results, we also conducted PPI analysis based on the human PPIs derived from the database of STRING (Szklarczyk et al., 2015) on November 7, 2017. The human PPIN was built based on the PPIs with median confidence (combined score equal to or greater than 0.4). It included 710,188 PPIs and 17,487 proteins which could be mapped to NCBI gene ids. Similar to those mentioned above, the viral receptor protein was observed to have far more interaction partners and higher betweenness than other proteins in the human PPIN (Figure S3A&C).

Evolutionary analysis

To identify the homolog of the mammalian virus receptor in other mammal species, the protein sequence of each viral receptor was searched against the database of mammalian protein sequences, which were downloaded from NCBI RefSeq database (Pruitt et al., 2012) on October 10th, 2017, with the help of BLAST (version 2.6.0) (Altschul et al., 1990). Analysis showed that in the database of mammalian protein sequences, there were 108 mammal species which were richly annotated and had far more protein sequences than other mammal species (Table S5). Therefore, only these 108 mammal species were considered in the evolutionary analysis. Based on the results of BLAST, the homolog for the viral receptor was defined as the hit with

E-value small than 1E-10, coverage equal to or greater than 80% and sequence identity equal to or greater than 30%. Only the closest homolog, i.e., the best hit, in each mammal species was used for further analysis. For measuring the conservation level of viral receptors, two indicators were used. The first indicator was the number of mammal species with homolog of the viral receptor in 108 mammal species. The other indicator was the average of the pairwise sequence identities between the viral receptor and its homologs in 108 mammal species. For comparison, 1000 human protein sequences were randomly selected from the NCBI RefSeq database (Table S4). Similar methods as above were utilized to calculate the indicators of conservation level for these proteins.

For analysis of co-evolution between viral receptors, firstly for each viral receptor, a phylogenetic tree was built based on the protein sequences of the receptor and its homologs in 108 mammal species with the help of phylip (version 3.68) (Felsenstein, 1989). The neighbor-joining method was used with the default parameter. Then, the genetic distances between the viral receptor and their homologs were extracted from the phylogenetic tree with a perl script. Finally, for a pair of viral receptors, the spearman correlation coefficient (SCC) was calculated between the pairwise genetic distances of viral receptors and their homologs, which was used to measure the extent of co-evolution between this pair of viral receptors.

The set of housekeeping gene in human was adapted from the work of Eisenberg et al (Eisenberg and Levanon, 2013). A total of 3804 genes were identified as the housekeeping gene.

Analysis of the tissue-specific expression of human viral receptors

The expression level for human viral receptors and other human genes in 32 human tissues were derived from the database of Expression Atlas (Petryszak et al., 2016). When comparing the expression level of human viral receptors and other set of human genes in 32 human tissues, to reduce the influence of extreme values, the median instead of the mean of the expression values was used to measure the average expression value of a gene set in a tissue.

The SCCs between the expression values of viral receptors in 32 tissues were calculated to measure the extent of co-expression between viral receptors.

Statistical analysis

All the statistical analysis was conducted in R (version 3.2.5) (Team, 2016). The wilcoxon rank-sum test was conducted with the function of *wilcox.test()*.

SUPPLEMENTAL TABLES

All the supplementary tables are provided in the Excel file.

Table S1. Related to Figure 1. The database of mammal virus-host receptor interaction.

Table S2. Related to Figure 2. Blast hit of the non-human mammal viral receptor against the human viral receptors.

Table S3. Related to Figure 2. All the significantly enriched GO terms and KEGG

pathways (adjusted p-values < 0.01) for the human viral receptor gene and those which directly interacted with the human viral receptor gene in the human protein-protein interaction network.

Table S4. Related to Figure S4. Conservation analysis for 1000 human proteins randomly selected from human proteome in the NCBI RefSeq database.

Table S5. Related to Figure S4. List of 108 mammal species which had far more protein sequences than other mammal species in NCBI RefSeq database

ACKNOWLEDGEMENTS

This work was supported by the National Key Plan for Scientific Research and Development of China (2016YFD0500300 and 2016YFC1200200), the National Natural Science Foundation of China (31500126, 31671371, 81730064 and 81571985), National Science and Technology Major Project (2017ZX10202201) and the Chinese Academy of Medical Sciences (2016-I2M-1-005). The authors would like to thank Pro. Xiangjun Du in Sun Yat-sen University for helpful suggestions.

The authors have declared that no competing interests exist.

AUTHOR CONTRIBUTIONS

HZ, TJ and YP conceived and designed the study. ZZ and ZZZ did the computational analysis. WC, ZC, ZT and BX compiled the database of mammalian viruses and their related receptors, and the tissue tropism of viruses. AW and XYG directed the computational analysis about the tissue and host specificity of viruses. YP and ZZ wrote the paper. AW, XYG, XHG, ZT, ZX, TJ and HZ reviewed and edited the manuscript. All authors read and approved the manuscript.

REFERENCES

Agarwala, R., Barrett, T., Beck, J., Benson, D.A., Bollin, C., Bolton, E., Bourexis, D., Brister, J.R., Bryant, S.H., Lanese, K., *et al.* (2016). Database resources of the National Center for Biotechnology Information. Nucleic Acids Res *44*, D7-D19.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. Journal of molecular biology *215*, 403-410.

Baranowski, E., Ruiz-Jarabo, C.M., and Domingo, E. (2001). Evolution of cell recognition by viruses. Science *292*, 1102-1105.

Bateman, A., Martin, M.J., O'Donovan, C., Magrane, M., Alpi, E., Antunes, R., Bely, B., Bingley, M., Bonilla, C., Britto, R., *et al.* (2017). UniProt: the universal protein knowledgebase. Nucleic Acids Res *45*, D158-D169.

Bernstein, H.J. (2017). RasMol 2.7.5. Available at http://www.openrasmol.org/.

Breban, R., Riou, J., and Fontanet, A. (2013). Interhuman transmissibility of Middle East respiratory syndrome coronavirus: estimation of pandemic risk. Lancet *382*, 694-699.

Casasnovas, J.M. (2013). Virus-Receptor Interactions and Receptor-Mediated Virus Entry into Host Cells, M.G. Mateu, ed. (Springer).

Corfield, A. (2017). Eukaryotic protein glycosylation: a primer for histochemists and cell biologists. Histochemistry and cell biology *147*, 119-147.

Dimitrov, D.S. (2004). Virus entry: Molecular mechanisms and biomedical applications. Nat Rev Microbiol *2*, 109-122.

Drayman, N., Glick, Y., Ben-nun-shaul, O., Zer, H., Zlotnick, A., Gerber, D., Schueler-Furman, O., and Oppenheim, A. (2013). Pathogens use structural mimicry of native host ligands as a mechanism for host receptor engagement. Cell host & microbe *14*, 63-73.

Eisenberg, E., and Levanon, E.Y. (2013). Human housekeeping genes, revisited. Trends Genet *29*, 569-574.

Felsenstein, J. (1989). PHYLIP - Phylogeny Inference Package (Version 3.2). Cladistics *5*, 3.

G, C., and T, N. (2006). The igraph software package for complex network research, InterJournal, Complex Systems 1695.

Ge, X.Y., Li, J.L., Yang, X.L., Chmura, A.A., Zhu, G.J., Epstein, J.H., Mazet, J.K., Hu, B., Zhang, W., Peng, C., *et al.* (2013). Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. Nature *503*, 535-+.

Geoghegan, J.L., Senior, A.M., Di Giallonardo, F., and Holmes, E.C. (2016). Virological factors that increase the transmissibility of emerging human viruses. Proceedings of the National Academy of Sciences of the United States of America *113*, 4170-4175.

Grove, J., and Marsh, M. (2011). The cell biology of receptor-mediated virus entry. The Journal of cell biology *195*, 1071-1082.

Hofmann, K., and Stoffel, W. (2017). TMpred: Prediction of Transmembrane Regions and Orientation. Available at https://embnet.vital-it.ch/software/TMPRED_form.html.

Isa, P., Arias, C.F., and Lopez, S. (2006). Role of sialic acids in rotavirus infection. Glycoconjugate journal *23*, 27-37.

Krupovic, M., and Koonin, E.V. (2017). Multiple origins of viral capsid proteins from cellular ancestors. Proceedings of the National Academy of Sciences of the United States of America *114*, E2401-E2410.

Li, F. (2015a). Receptor Recognition Mechanisms of Coronaviruses: a Decade of Structural Studies. Journal of virology *89*, 1954-1964.

Li, F. (2016). Structure, Function, and Evolution of Coronavirus Spike Proteins. Annu Rev Virol *3*, 237-261.

Li, W. (2015b). The hepatitis B virus receptor. Annual review of cell and developmental biology *31*, 125-147.

Maganga, G.D., Kapetshi, J., Berthet, N., Ilunga, B.K., Kabange, F., Kingebeni, P.M., Mondonge, V., Muyembe, J.J.T., Bertherat, E., Briand, S., *et al.* (2014). Ebola Virus Disease in the Democratic Republic of Congo. New Engl J Med *371*, 2083-2091.

Maginnis, M.S., Haley, S.A., Gee, G.V., and Atwood, W.J. (2010). Role of N-linked glycosylation of the 5-HT2A receptor in JC virus infection. Journal of virology *84*, 9677-9684.

Marija Backovic, F.A.R. (2012). Virus entry: old viruses, new receptors. Current opinion in virology *2*, 10.

Masson, P., Hulo, C., De Castro, E., Bitter, H., Gruenbaum, L., Essioux, L., Bougueleret, L., Xenarios, I., and Le Mercier, P. (2013). ViralZone: recent updates to the virus knowledge resource. Nucleic Acids Res *41*, D579-D583.

Mazzon, M., and Mercer, J. (2014). Lipid interactions during virus entry and infection. Cellular microbiology *16*, 1493-1502.

Menche, J., Sharma, A., Kitsak, M., Ghiassian, S.D., Vidal, M., Loscalzo, J., and Barabasi, A.L. (2015). Uncovering disease-disease relationships through the incomplete interactome. Science *347*.

Mihara, T., Nishimura, Y., Shimizu, Y., Nishiyama, H., Yoshikawa, G., Uehara, H., Hingamp, P., Goto, S., and Ogata, H. (2016). Linking Virus Genomes with Host Taxonomy. Viruses-Basel *8*.

Mlakar, J., Korva, M., Tul, N., Popovic, M., Poljsak-Prijatelj, M., Mraz, J., Kolenc, M., Rus, K.R., Vipotnik, T.V., Vodusek, V.F., *et al.* (2016). Zika Virus Associated with Microcephaly. New Engl J Med *374*, 951-958.

Olival, K.J., Hosseini, P.R., Zambrana-Torrelio, C., Ross, N., Bogich, T.L., and Daszak, P. (2017). Host and viral traits predict zoonotic spillover from mammals. Nature *546*, 646-+.

Paez-Espino, D., Eloe-Fadrosh, E.A., Pavlopoulos, G.A., Thomas, A.D., Huntemann, M., Mikhailova, N., Rubin, E., Ivanova, N.N., and Kyrpides, N.C. (2016). Uncovering Earth's virome. Nature *536*, 425-+.

Peng, W.J., de Vries, R.P., Grant, O.C., Thompson, A.J., McBride, R., Tsogtbaatar, B., Lee, P.S., Razi, N., Wilson, I.A., Woods, R.J., *et al.* (2017). Recent H3N2 Viruses Have Evolved Specificity for Extended, Branched Human-type Receptors, Conferring Potential for Increased Avidity. Cell host & microbe *21*, 23-34.

Petryszak, R., Keays, M., Tang, Y.A., Fonseca, N.A., Barrera, E., Burdett, T., Füllgrabe, A., Fuentes, A.M.-P., Jupp, S., Koskinen, S., *et al.* (2016). Expression Atlas update—an integrated database of gene and protein expression in humans, animals and plants. Nucleic Acids Res *44*, D746-D752.

Pillay, S., Meyer, N.L., Puschnik, A.S., Davulcu, O., Diep, J., Ishikawa, Y., Jae, L.T., Wosen, J.E., Nagamine, C.M., Chapman, M.S., *et al.* (2016). An essential receptor for adeno-associated virus infection. Nature *530*, 108-112.

Pruitt, K.D., Tatusova, T., Brown, G.R., and Maglott, D.R. (2012). NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. Nucleic Acids Res *40*, D130-D135.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Roy, A., Kucukural, A., and Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc *5*, 725-738.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome research *13*, 2498-2504.

Sharp, P.M., and Hahn, B.H. (2011). Origins of HIV and the AIDS Pandemic. Csh Perspect Med *1*.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., *et al.* (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res *43*, D447-D452.

Wang, J.H. (2002). Protein recognition by cell surface receptors: physiological receptors versus virus interactions. Trends in biochemical sciences *27*, 122-126.

Yan, H., Zhong, G., Xu, G., He, W., Jing, Z., Gao, Z., Huang, Y., Qi, Y., Peng, B., Wang, H., *et al.* (2012). Sodium taurocholate cotransporting polypeptide is a functional receptor for human hepatitis B and D virus. eLife *1*, e00049.

Yu, G.C., Wang, L.G., Han, Y.Y., and He, Q.Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. Omics *16*, 284-287.

Supplementary Figures

Figure S1. Related to Figure 1. Description of mammalian virus-host receptors. **(A)** Host composition for the viral receptor. The numbers in parentheses refer to the number of viral receptor for the host. (B) Distribution of the number of receptors used by a virus. (C) Distribution of the number of viruses using the same viral receptor.

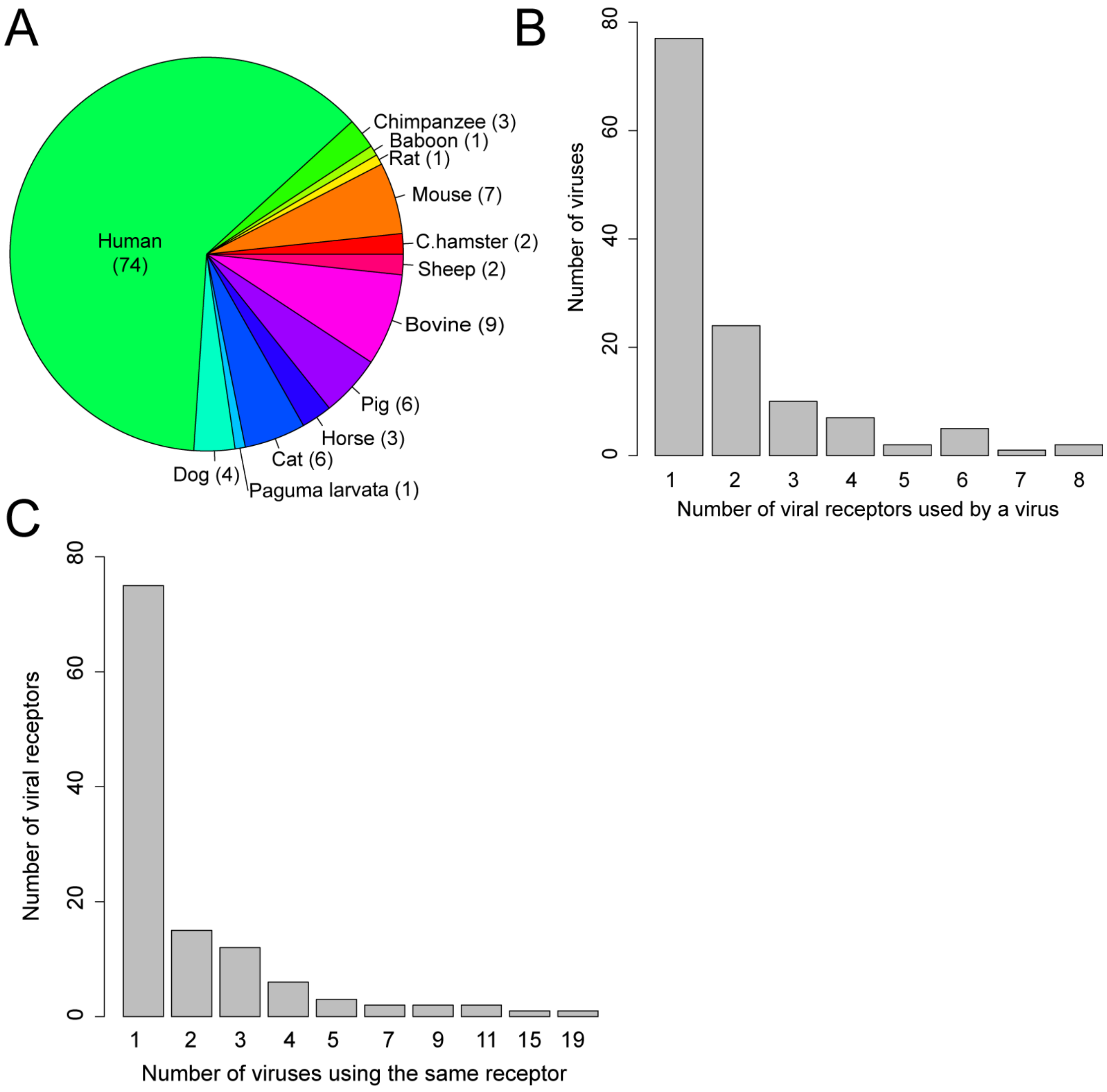


Figure S2. Related to Figure 2. Structural analysis of the mammalian virus receptors. (A) Distribution of the number of transmembrane alpha helix within a viral receptor protein for the mammalian viral receptor. (B) Comparing the number of Pfam domains within a protein for the set of human proteins, human membrane proteins, human cell membrane proteins, human viral receptors and mammalian viral receptors. For clarity, all the outliers greater than 6 were removed from the figure. "***", p-value < 0.001. (C) Top ten Pfam families observed in the mammalian viral receptor proteins. (D) Comparing O-glycosylation levels between mammal viral receptors, human viral receptors, human cell membrane proteins, human membrane proteins and all human proteins. There were no significant differences between them (p-value > 0.1 in the Wilcoxon rank-sum test).

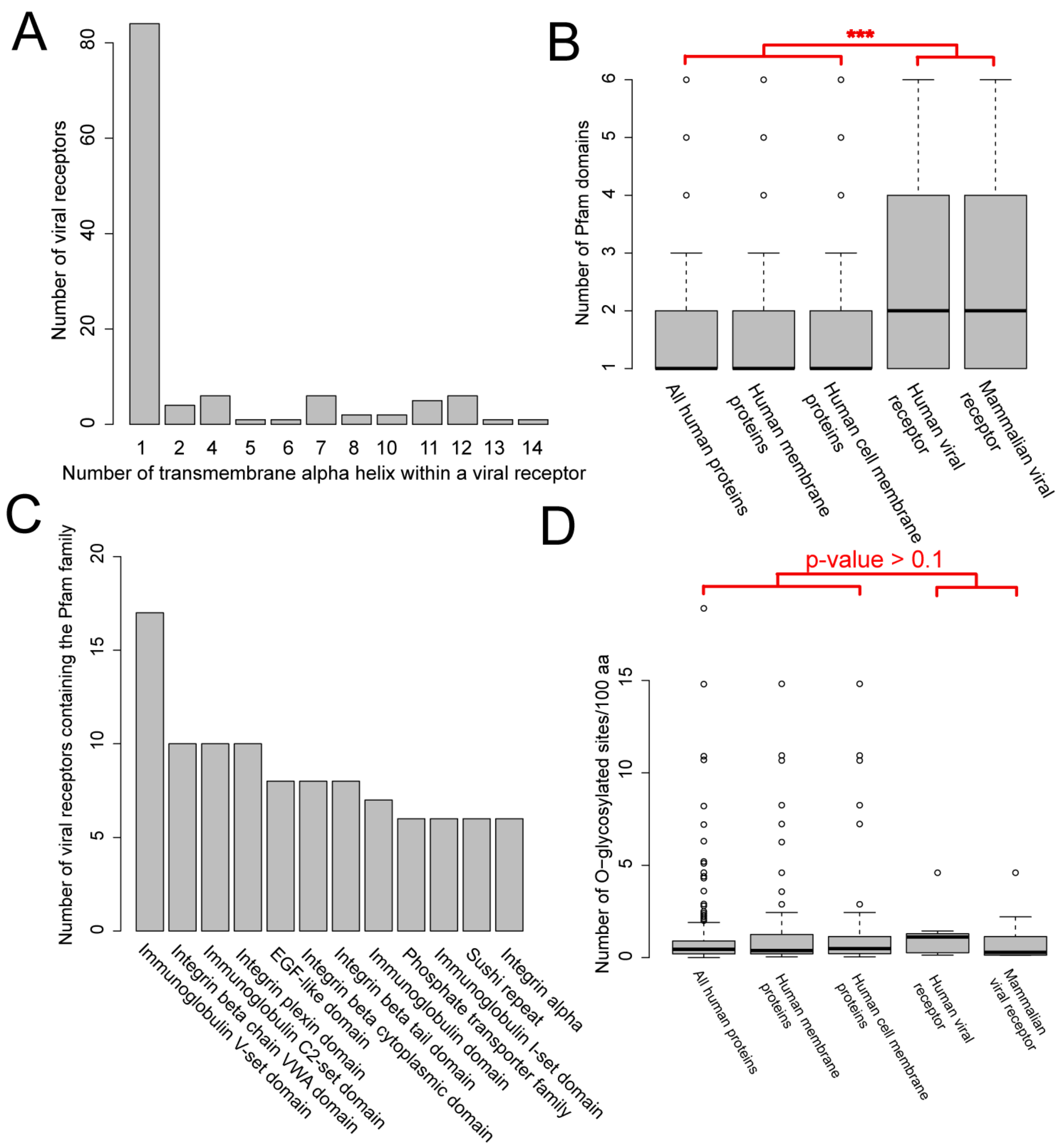

A
Number of viral receptors
0
20
40
60
80
1
2
4
5
6
7
8
10
11
12
13
14
Number of transmembrane alpha helix within a viral receptor
B
***
Number of Pfam domains
1
2
3
4
5
6
All human proteins
Human membrane proteins
Human cell membrane proteins
Human viral receptor
Mammalian viral receptor
C
Number of viral receptors containing the Pfam family
0
5
10
15
20
Immunoglobulin V-set domain
Integrin beta chain VWA domain
Immunoglobulin C2-set domain
Integrin plexin domain
EGF-like domain
Integrin beta cytoplasmic domain
Integrin beta tail domain
Immunoglobulin domain
Phosphate transporter family
Immunoglobulin I-set domain
Sushi repeat
Integrin alpha
D
p-value > 0.1
Number of O−glycosylated sites/100 aa
0
5
10
15
All human proteins
Human membrane proteins
Human cell membrane proteins
Human viral receptor
Mammalian viral receptor

Figure S3. Related to Figure 2. Protein-protein interaction analysis of human virus receptors. (A) Comparing the degree of proteins in the set of human viral receptors, human cell membrane proteins, human membrane proteins and all human proteins based on the human PPI network derived from the database of STRING. "***", p-value < 0.001 in the Wilcoxon rank-sum test. (B) Comparing the betweenness of proteins in the set of human viral receptors, human cell membrane proteins, human membrane proteins and all human proteins based on the human PPI network derived from the work of Menche et al. "***", p-value < 0.001 in the Wilcoxon rank-sum test. (C) Comparing the betweenness of proteins in the set of human viral receptors, human cell membrane proteins, human membrane proteins and all human proteins based on the human PPI network derived from the database of STRING. "***", p-value < 0.001 in the Wilcoxon rank-sum test. (D) The partial human PPI network composed of only human viral receptors. The gene symbol for each node was presented. The node in blue refers to those with self-interactions.

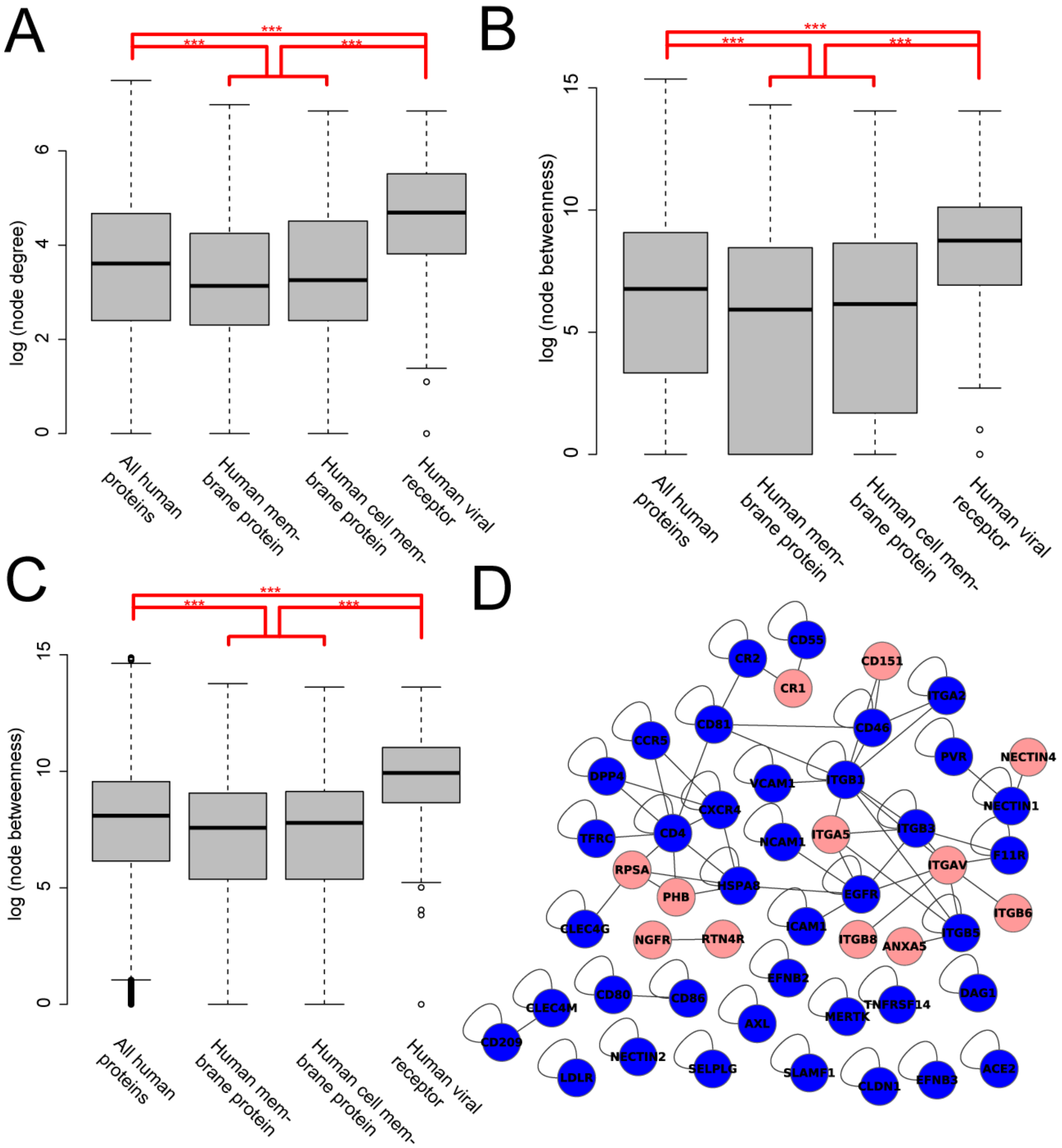
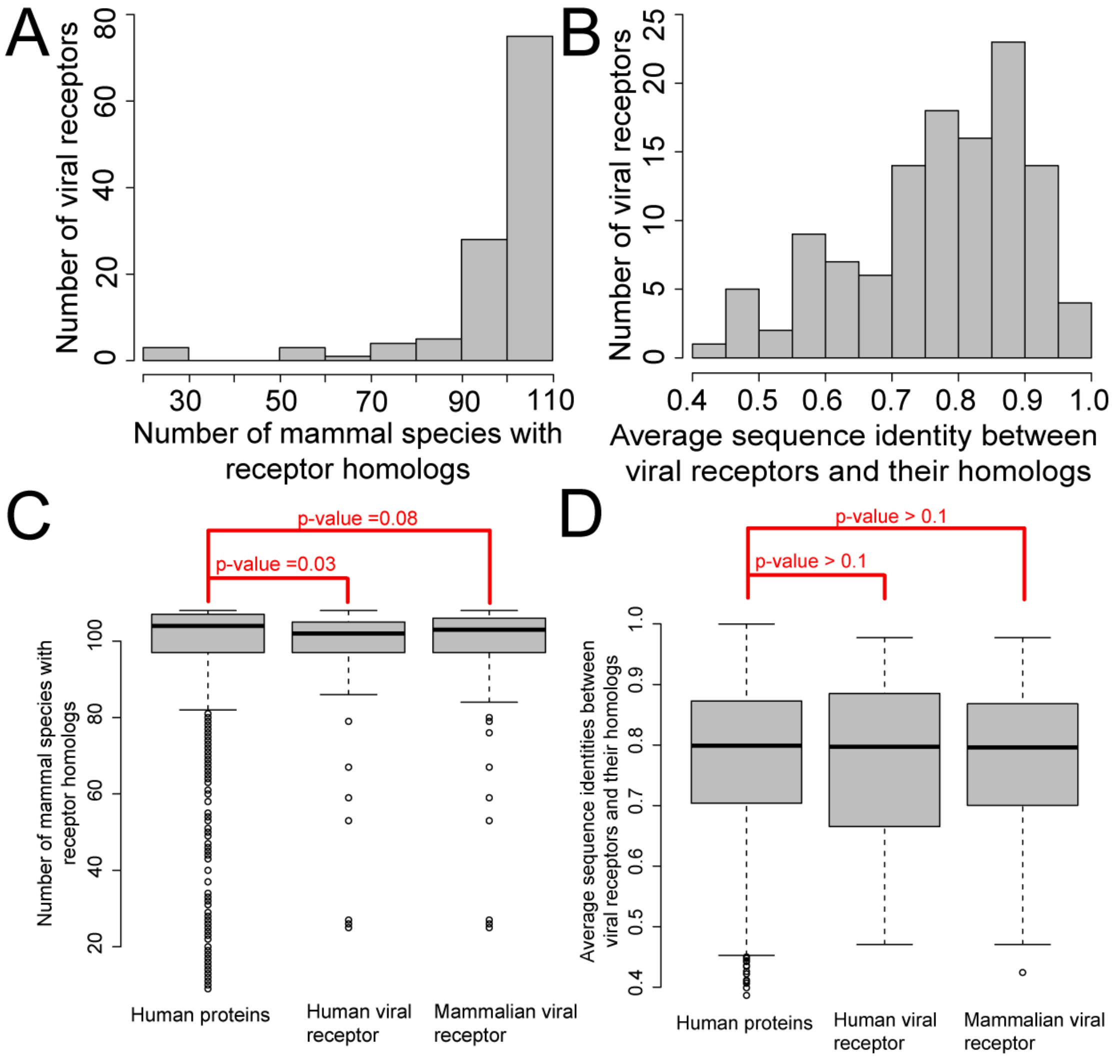

A
***
***
***
log (node degree)
0
2
4
6
All human proteins
Human membrane protein
Human cell membrane protein
Human viral receptor
B
***
***
***
log (node betweenness)
0
5
10
15
All human proteins
Human membrane protein
Human cell membrane protein
Human viral receptor
C
***
***
***
log (node betweenness)
0
5
10
15
All human proteins
Human membrane protein
Human cell membrane protein
Human viral receptor
D
CD55
CR2
CR1
CD151
ITGA2
CD81
CD46
CCR5
PVR
NECTIN4
DPP4
VCAM1
ITGB1
NECTIN1
CXCR4
TFRC
CD4
NCAM1
ITGA5
ITGB3
F11R
RPSA
ITGAV
PHB
HSPA8
EGFR
ITGB6
CLEC4G
ICAM1
NGFR
RTN4R
ITGB8
ANXA5
ITGB5
EFNB2
CD80
CD86
DAG1
CLEC4M
TNFRSF14
AXL
MERTK
CD209
NECTIN2
SELPLG
SLAMF1
ACE2
LDLR
CLDN1
EFNB3

Figure S4. Related to Figure 4. Evolutionary analysis of the mammalian virus receptor. (A) Distribution of the number of mammal species with homologs to the viral receptor in 108 mammal species. (B) Distribution of the average sequence identities between viral receptors and their homologs in mammal species. (C) Comparing the number of species among 108 mammal species which had homologs to the viral receptor for the set of mammalian viral receptors, human viral receptors and 1000 human proteins randomly selected (see Table S4). (D) Comparing the average sequence identities between viral receptors and their homologs in 108 mammal species for the set of mammal viral receptors, human viral receptors and 1000 human proteins randomly selected (see Table S4).


A
Number of viral receptors
0
20
40
60
80
30
50
70
90
110
Number of mammal species with receptor homologs
B
Number of viral receptors
0
5
10
15
20
25
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Average sequence identity between viral receptors and their homologs
C
p-value =0.08
p-value =0.03
Number of mammal species with receptor homologs
20
40
60
80
100
Human proteins
Human viral receptor
Mammalian viral receptor
D
p-value > 0.1
p-value > 0.1
Average sequence identities between viral receptors and their homologs
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Human proteins
Human viral receptor
Mammalian viral receptor