Prediction of receptorome for human-infecting virome

Zheng Zhang[1,#], Sifan Ye[1,#], Aiping Wu[2], Taijiao Jiang[2],Yousong Peng[1,*]

[1] College of Biology, Hunan Provincial Key Laboratory of Medical Virology, Hunan University, Changsha 410082, China

[2] Center for Systems Medicine, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100005, Suzhou Institute of Systems Medicine, Suzhou, Jiangsu 215123, China;

[#] These authors contributed equally to this work

* Corresponding author: Yousong Peng, pys2013@hnu.edu.cn

Abstract

The virus receptor is key for viral infection of host cells. Identification of the virus receptor is still challenging at present. Our previous study has shown that human virus receptor proteins have some unique features including high level of N-glycosylation, high number of interaction partners and high expressions. Here, we built a random-forest model to identify human virus receptorome from human cell membrane proteins with accepted accuracy based on combination of unique features of human virus receptors and protein sequences. A total of 729 human cell membrane proteins were predicted to constitute the receptorome for the human-infecting virome. Combination of the random-forest model with protein-protein interactions between human and viruses predicted in previous studies further predicted receptors for 693 human-infecting viruses, such as the Enterovirus, Norovirus and West Nile virus. Finally, we predicted the candidate alternative receptors for the 2019-nCoV. The study

would greatly facilitate identification of receptors for human-infecting virome.

Introduction

Receptor-binding is the first step for viral infection of host cells. Proteins are considered to be better receptors for viruses due to their higher binding affinity and specificity than carbohydrate and lipid (Baranowski et al., 2001; Casasnovas, 2013; Dimitrov, 2004; Li, 2015; Wang, 2002). Lots of human virus receptors has been identified. The selection of proteins by viruses as receptors is not a random process. 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 (Dimitrov, 2004; Wang, 2002). Based on a collection of 119 mammalian virus receptors, our recent study further showed that human virus protein receptors have higher level of N-glycosylation, higher number of interaction partners in the human protein-protein interaction (PPI) network and high expressions in 32 common human tissues compared to other cell membrane proteins (Zhang et al., 2019). These studies could facilitate identification of human virus receptors.

Identifying receptors for viruses is challenging. Currently, several experimental approaches have been developed in identifying virus receptors. The first approach is to select candidate membrane proteins that bind to the virus receptor-binding proteins (RBPs) by affinity purification and mass spectroscopy (Free et al., 2009; Ryu, 2016). The second approach is to select candidate membrane proteins to which the monoclonal antibodies specific could block the virus entry (Minor et al., 1984; Ryu, 2016). The third approach is to identify the virus receptor by functional cloning selection based on the cDNA expression library (Ryu, 2016). The protein which could enable viral infection of the non-susceptible cell after transfection would be considered as receptor candidate. However, identifying virus receptor is still time-consuming and difficult at

present. Besides, limitations on scalability have hampered large-scale identification of viral receptors. The development of computational methods for identifying the human virus receptors is in great need.

Previous studies have developed computational models for predicting PPIs between viruses and hosts which could help identification of virus receptors (Lasso et al., 2019; Yan et al., 2019). For example, Lasso et al. developed an in silico computational framework (P-HIPSTer) that employs structural information to predict more than 280,000 PPIs between 1,001 human-infecting viruses and humans, and made a series of new findings about human-virus interactions (Lasso et al., 2019). Numerous PPIs between viral RBPs and human cell membrane proteins could be used to identify virus receptors. Here, we developed a computational model to predict the receptorome of the human-infecting virome based on the features of human virus receptors and protein sequences. Combination of the model with Lasso's work further predicted receptors for 693 human-infecting viruses. The study would greatly facilitate identification of human virus receptors.

Materials and Methods

Human viral receptors, human cell membrane proteins and human membrane proteins

A total of 91 human virus protein receptors were obtained from the viralReceptor database (available at http://www.computationalbiology.cn:5000/viralReceptor) developed in our previous study (Zhang et al., 2019). Human cell membrane proteins and human membrane proteins were obtained from the UniProtKB/Swiss-Prot database on November 24th, 2017. The human proteins with the words of "cell membrane" and "cell surface" in the field of "Subcellular location" were considered to be human cell

membrane proteins. The human proteins with the words of “membrane” in the field of “Subcellular location” were considered to be human membrane proteins. A total of 2,161 human cell membrane proteins and 5,187 human membrane proteins were obtained.

The protein features and protein sequence features

The N-glycosylation sites of human proteins were obtained from the UniprotKB/Swiss-Prot database. For those proteins without annotation of the N-glycosylation site in the UniprotKB/Swiss-Prot database, the N-glycosylation sites were predicted with NetNGlyc 1.0 (available at http://www.cbs.dtu.dk/services/NetNGlyc/) (Gupta et al., 2004). The level of N-glycosylation for a protein was defined as the number of N-glycosylation sites per hundred amino acids.

To calculate the node degree of human proteins in the human PPI network, firstly, the human PPIs with combined scores greater than 400 were extracted from the STRING database (version 10.5) (Szklarczyk et al., 2015) and formed the human PPI network. Then, the node degree was calculated with the functions of *degree* in the R package igraph (version 1.2.4.2) (Csardi and Nepusz, 2006).

The expression levels of human genes in 32 common human tissues were obtained from the Expression Atlas database (Petryszak et al., 2016) on February 6th, 2018. Because there were strong correlations between the expressions of genes in different tissues, the principal component analysis (PCA) method was used to reduce the correlations with the function of *PCA* in the package scikit-learn (version 0.21.3) (Pedregosa et al., 2011) in Python (version 3.6.7). Only the first principal component was used to measure the expression level of human genes, which explained 95% of the

total variance.

The frequencies of k-mers with one or two amino acids were calculated for each human protein with a Python script.

Machine learning modeling

To distinguish human virus receptors from other human cell membrane proteins with machine learning models, the human virus receptors were taken as positive samples, and the human membrane proteins were taken as negative samples after excluding the human cell membrane proteins since the human cell membrane proteins possibly contain virus receptors unidentified yet.

Because not all human proteins were observed in the human PPI network used or had observed expressions in the data used, only the human proteins with all three protein features, i.e., the N-glycosylation level, node degree and expression in common human tissues, were used in the modeling. Besides, the sequence redundancy in both human virus receptor proteins and human membrane proteins were removed using CD-HIT (version 4.8.1) (Fu et al., 2012) at 70% identity level. Finally, a total of 88 human virus receptors and 1,593 human membrane proteins was used in the machine learning modeling.

The random-forest (RF) model is an ensemble machine learning technique using multiple decision trees and can handle data with high variance and high bias, while the risk of over-fitting can be significantly reduced by averaging multiple trees. Therefore, we chose the RF model to distinguish human virus receptors from other human cell membrane proteins. Because the number of positive samples, i.e., human viral

receptors, was much smaller than that of negative samples, i.e., human membrane proteins, the function of BalanceRandomForest in the package imbalanced-learn (version 0.5.0) (Chen et al., 2004) in Python was used to deal with the imbalanced positive and negative samples with the parameter of n_estimators set to be 100.

Five times of five-fold cross-validations were used to evaluate the predictive performances of the RF model with the function of StratifiedKFold in the package scikit-learn in Python. The predictive performances of the RF model were evaluated by the area under receiver operating characteristics curve (AUC), accuracy, sensitivity and specificity.

Validation of the RF model in ranking receptor candidates

The predicted interactions between human-infecting viruses and human proteins were obtained from the database of P-HIPSTer (available at http://phipster.org/) (Lasso et al., 2019) on November 1st, 2019. A total of 9,395 pairs of interactions between viral RBPs and human cell membrane proteins with the likelihood ratio (LR) $\geq$100 were extracted for further analysis, which included 718 viral RBPs and 314 human cell membrane proteins. The RBPs of human-infecting viruses were compiled from three sources: the ViralZone database (Masson et al., 2012), the UniprotKB database during which viral proteins annotated with GO terms “viral entry into host cell” or “virion attachment the host cell”, and the literatures related to viral RBPs. Viruses of the same viral family were supposed to use the same RBP. For example, all coronaviruses were supposed to use the spike protein as RBP.

To evaluate the ability of the RF model in helping identify virus receptors, 25 pairs of experimentally-validated interactions between viral RBPs and receptors, and the predicted PPIs between these viral RBPs and human cell membrane proteins, were

extracted from the P-HIPSTer database. For each viral RBP, the predicted RBP-interacting human cell membrane proteins were ranked by either the LR provided in Lasso's work, or the predicted score provided by the RF model. Then, the ranks of real receptors were analyzed. The rank percentage of the real receptor was calculated by dividing the rank by the number of RBP-interacting human cell membrane proteins.

When ranking the RBP-interacting proteins by the RF model, high sequence similarity between the RBP-interacting proteins and human proteins in the modeling (both the positive and negative samples) may over-estimate the ability of the RF model. To reduce such effect, for each pair of viral RBP and receptor, the predicted RBP-interacting human proteins were clustered with human proteins used in the modeling using CD-HIT at 50% identity level. All proteins which were clustered with RBP-interacting proteins were excluded in the modeling.

Results

Development of random-forest models for predicting receptorome of human-infecting virome

Our previous studies have shown that human virus protein receptors have high level of N-glycosylation, high number of interaction partners in the human PPI network, and high expressions in 32 common human tissues (Zhang et al., 2019). To identify the potential receptors of human-infecting virome, we firstly built a RF model to distinguish the human virus receptor proteins from other human membrane proteins based on these protein features. The RF model built based on individual protein feature achieved an AUC ranging from 0.50 to 0.66 in five-fold cross-validations (Table 1). Combination of all three features improved the RF model much, with both the AUC and prediction accuracy equaling to 0.75 (Table 1).

For comparison, we also developed RF models to distinguish the human virus from other human membrane proteins based on protein sequences. The frequencies of k-mers with length of one and two amino acids were used as features in the modeling. The RF model based on combination of 20 amino-acid frequencies achieved an AUC 0.68. When using the frequencies of k-mers with two amino acids as features, only the top N (N=1 to 400) most important features were used in the modeling to reduce over-fitting. Figure S1 showed that the AUC of the RF models increased rapidly as increase of N from 1 to 10. Then, it increased slowly or even decreased when more ten features were used. Therefore, only top ten features were used in the modeling to balance the model performance and complexity. The RF model based on top ten features achieved an AUC of 0.72.

To further improve the model for predicting receptorome of human-infecting virome, we incorporated the protein features and the top ten most important k-mers of two amino acids in the modeling. The RF model achieved an AUC of 0.77. The prediction accuracy, sensitivity and specificity of the model equaled 0.80, 0.73 and 0.80, respectively (Table 1). The model combining both the protein features and k-mers was used for further analysis.

Table 1. The predictive performances of RF models using different sets of features. N-gly, N-glycosylation; PPI, node degree in human PPI network; Expression, expressions in 32 human tissues. Acc, accuracy; Sen, Sensitivity; Spe, Specificity.

Model with different sets of features	**Acc**	**Sen**	**Spe**	**AUC**

N-gly	0.58	0.63	0.57	0.60
PPI	0.68	0.65	0.68	0.66
Expression	0.50	0.51	0.50	0.50
N-gly+PPI+Expression	0.75	0.75	0.75	0.75
Top 10 k-mers	0.73	0.71	0.73	0.72
N-gly+PPI+Expression + Top 10 k-mers	0.80	0.73	0.80	0.77

Prediction of receptorome for human-infecting virome

Based on the RF model, we predicted the receptorome from human cell membrane proteins. A score ranging from 0 to 1 was assigned to each human cell membrane protein. Proteins with high scores are more likely to be virus receptors. A total of 729 proteins with scores greater than 0.5 were considered to constitute the receptorome of human-infecting virome. All human cell membrane proteins and the relevant scores were provided in Table S1.

Prediction of virus-receptor interactions for human-infecting viruses

Then, we further investigated the prediction of virus-receptor interactions. In previous study, Lasso et al. predicted 282,528 pairs of PPIs between human and 1,001 human-infecting viruses (Lasso et al., 2019). Based on their study, 9,395 pairs of PPIs between 718 viral RBPs from 693 human-infecting viruses, and 314 human cell membrane proteins were extracted for further analysis (see Table S2). A viral RBP was predicted to interact with 1 to 65 human cell membrane proteins, with a median of 10. For each viral RBP, the RBP-interacting cell membrane proteins were ranked by the score provided by the RF model to select the most likely receptor (Table S2).

To validate the ranking by the RF model, 25 pairs of experimentally-validated interactions between viral RBPs and receptors were extracted. For each pair of these interactions, the rank of the real receptor among the predicted RBP-interacting proteins was obtained, and the related rank percentage was calculated (Materials and Methods). Seven real receptors were ranked in top one by the RF model (Table 2). Besides, nearly 70% (17/25) of real receptors were ranked in top three. On average, the real receptors had a rank percentage of 0.21 among all RBP-interacting human cell membrane proteins, suggesting that the real receptors would be ranked in the top 21% of all candidates by the RF model.

For comparison, we also ranked the RBP-interacting proteins by the LR provided in Lasso's work. In 12 pairs of viral RBP-receptor interaction with LRs, no real receptors were ranked in top one and only two real receptors were ranked in top three when ranking RBP-interacting human cell membrane proteins by the LR. On average, the median rank percentage of real receptors was 0.43 when ranking by the LR, while that was 0.15 by the RF model (Table 2).

Table 2. The ranks of real virus receptors among the RBP-interacting human cell membrane proteins by LR and RF score. The median rank percentage of real virus receptors among RBP-interacting human cell membrane proteins, and the number of real virus receptors among top one, three and five ranks were summarized at the bottom. [#] The number in brackets referred to those when only considering the 12 pairs of viral RBP-receptor interaction with LRs. * No LR was provided in Lasso's work since there were resolved complex structures between the RBP and the receptor. MACV, Machupo mammarenavirus; HAdV-A, Human mastadenovirus A; HAdV-C, Human mastadenovirus C; HAdV-E, Human mastadenovirus E;HAdV-D, Human

mastadenovirus D; HSV-1, Human alphaherpesvirus 1; HSV-2, Human alphaherpesvirus 2; HIV-1, Human immunodeficiency virus 1.

Virus name	**RBP**	**Real viral receptor**	**Num of RBP-interacting proteins**	**Rank by LR**	**Rank by RF score**
SARS-CoV	S	ACE2	31	-*	24
MERS-CoV	S	DPP4	8	-	3
Echovirus E6	VP1	CD55	13	5	2
Echovirus E11	VP1	CD55	9	4	1
Echovirus E7	VP1	CD55	7	-	3
Echovirus E13	VP1	CD55	11	4	1
Echovirus E20	VP1	CD55	12	5	2
Echovirus E29	VP1	CD55	13	6	2
Echovirus E33	VP1	CD55	13	6	1
Enterovirus C	VP1	PVR	5	-	2
Hepacivirus C	E	EGFR	17	10	1
MACV	Glyco-protein	TFRC	2	-	1
Measles virus	H	NECTIN4	18	-	17
Measles virus	H	SLAMF1	18	2	5

Hendra virus	G	EFNB2	5	-	1
Nipah virus	G	EFNB2	5	-	1
HAdV-A	L5	CXADR	25	-	12
HAdV-C	L5	CXADR	5	4	3
HAdV-E	L5	CXADR	33	3	20
HAdV-D	L5	CXADR	28	4	17
HSV-1	US6	TNFRSF14	28	-	6
HSV-1	US6	NECTIN1	28	-	3
HSV-2	US6	NECTIN1	34	-	7
HSV-2	US6	TNFRSF14	34	23	3
HIV-1	env	CD4	21	-	2
Top 1				0	7 (4)[#]
Top 3				2	17 (9)[#]
Top 5				8	18 (10)[#]
Median rank percentage				0.43	0.21 (0.15)[#]

Prediction of candidate alternative receptors for 2019-nCoV

Previous studies have shown that the ACE2 protein which is the receptor of 2019-nCoV (Hoffmann et al., 2020; Zhou et al., 2020) had a low level of expression in the lung and the upper respiratory tract (Qi et al., 2020; Zhang et al., 2020). It is

possible that 2019-nCoV has alternative receptors. We investigated the prediction of the alternative receptors for 2019-nCoV. Lasso's study have predicted PPIs between 28 human cell membrane proteins which were members of receptorome of human-infecting viruses, and the spike protein of two coronaviruses, including Severe Acute Respiratory Syndrome-CoV and Middle East Respiratory Syndrome-CoV. We supposed that the 2019-nCoV is very likely to use these spike-interacting proteins as alternative receptors. These spike-interacting proteins were ranked by the score provided by the RF model (Figure 1). The expression levels of these spike-interacting proteins in 32 common human tissues were shown in Figure 1. Most of them had high expressions than ACE2 in the lung, such as APP, EZR, CD4 and so on.

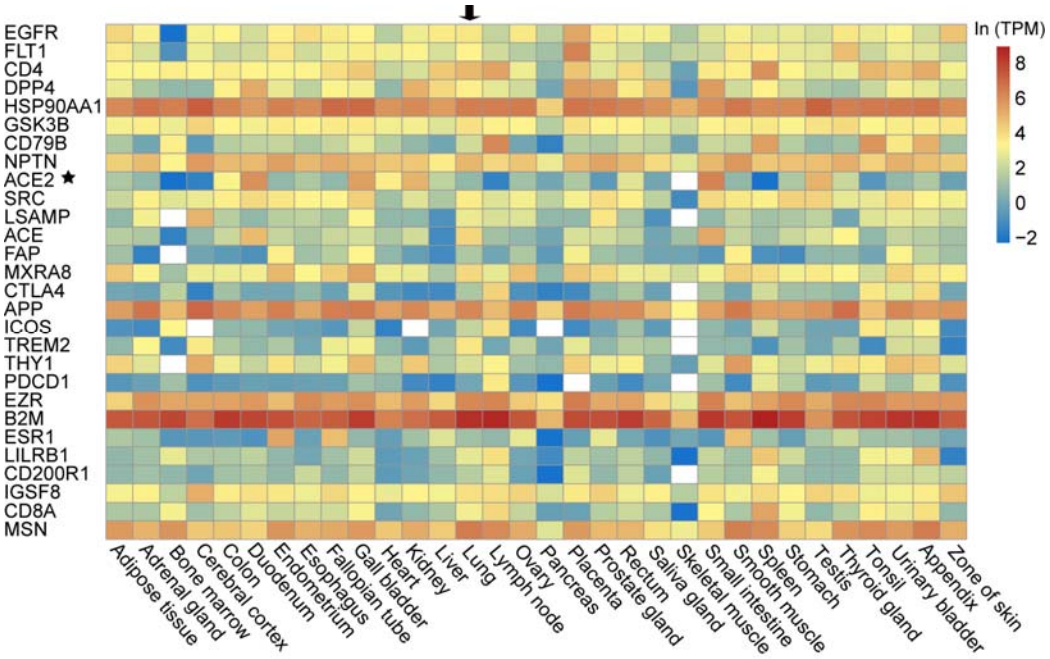

Figure 1. The predicted alternative receptors (in the left side) of 2019-nCoV and their expressions in 32 human tissues (in the bottom). The predicted alternative receptors were ranked by the RF score. The expression level was measured by transcripts per million (TPM) and was colored according to the legend in the top right. The white referred to no data available. The lung was highlighted by an arrow. The ACE2 was marked by an asterisk.

Discussion

Identification of receptors for human-infecting viruses is critical for understanding the interactions between viruses and human. Our previous studies have shown that human virus receptors have some unique features compared to other cell membrane proteins: high level of N-glycosylation, high number of interaction partners and high expressions. This study further built a RF model for identifying human virus receptors from human cell membrane proteins with accepted accuracy. Based on the RF model, the receptorome for human-infecting virome were predicted which included a total of 729 human cell membrane proteins. They could facilitate identification of human virus receptors.

Previous studies by Lasso et al. developed computational models for predicting PPIs between human-infecting viruses and humans (Lasso et al., 2019). A variable number of human cell membrane proteins were predicted to interact with viral RBPs. To further select the potential receptors for viruses, both the LR and the RF model could be used to rank the viral RBP-interacting human cell membrane proteins. The RF model was found to rank the real receptors better than the LR in a small validation dataset, suggesting that the RF model may perform better than the LR in selecting the real receptors from the predicted RBP-interacting human cell membrane proteins. Combination of the RF model and the RBP-interacting human cell membrane proteins predicted in Lasso's work enables prediction of receptors for 693 human-infecting viruses (Table S2). Further efforts are needed to validate these predicted receptors.

The 2019-nCoV is responsible for the on-going outbreak of COVID-19 (Wu et al., 2020; Zhou et al., 2020). Previous studies have shown that the virus used ACE2 as the receptor (Hoffmann et al., 2020; Zhou et al., 2020). However, several studies have

found that ACE2 had low expressions in the lung and the upper respiratory tract (Qi et al., 2020; Zhang et al., 2020). It is possibly that the 2019-nCoV may use co-receptor or alternative receptors to facilitate its infection. A previous study investigated the candidate co-receptors of 2019-nCoV, such as ANPEP, ENPEP and DPP4 (Qi et al., 2020). We further supposed that the 2019-nCoV possibly have alternative receptors. Based on the predicted PPIs between coronaviruses and humans in previous work, we predicted several candidate alternative receptors for 2019-nCoV. Most candidate alternative receptors had higher expressions than ACE2 in the lung. Nevertheless, more efforts are needed to validate these candidate alternative receptors.

There are some limitations to this study. Firstly, the number of human virus receptor proteins was much less than that of human membrane proteins in the modeling, which may hinder accurate modeling. The under-sampling method was used to deal with the imbalance problem. Secondly, the performance of the RF model was modest in discriminating the human virus receptor proteins from human membrane proteins. More efforts are needed to improve the model. Thirdly, the RF model can not be used to identify receptors for a specific human-infecting virus. Instead, it predicts the receptorome for the human-infecting virome. Combining the RF model with the model of PPI predictions such as Lasso's work can help identify virus-receptor interactions.

In conclusion, this study for the first time built a computational model for predicting the receptorome of the human-infecting virome. It could facilitate identification of human virus receptors in either computational or experimental studies.

Acknowledgements

This work was supported by the National Key Plan for Scientific Research and

Development of China (2016YFD0500300), Hunan Provincial Natural Science Foundation of China (2018JJ3039), and the Chinese Academy of Medical Sciences (2016-I2M-1-005).

The authors have declared that no competing interests exist.

References

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

Casasnovas, J.M. (2013). Virus-receptor interactions and receptor-mediated virus entry into host cells. In Structure and Physics of Viruses (Springer), pp. 441-466.

Chen, C., Liaw, A., and Breiman, L. (2004). Using random forest to learn imbalanced data. University of California, Berkeley *110*, 24.

Csardi, G., and Nepusz, T. (2006). The igraph software package for complex network research. InterJournal, complex systems *1695*, 1-9.

Dimitrov, D.S. (2004). Virus entry: molecular mechanisms and biomedical applications. Nature Reviews Microbiology *2*, 109-122.

Free, R.B., Hazelwood, L.A., and Sibley, D.R. (2009). Identifying novel protein‐protein interactions using co‐immunoprecipitation and mass spectroscopy. Current protocols in neuroscience *46*, 5.28. 21-25.28. 14.

Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics *28*, 3150-3152.

Gupta, R., Jung, E., and Brunak, S. (2004). Prediction of N-glycosylation sites in human proteins.

Hoffmann, M., Kleine-Weber, H., Krueger, N., Mueller, M.A., Drosten, C., and Pöhlmann, S. (2020). The novel coronavirus 2019 (2019-nCoV) uses the SARS-coronavirus receptor ACE2 and the cellular protease TMPRSS2 for entry into target cells. bioRxiv.

Lasso, G., Mayer, S.V., Winkelmann, E.R., Chu, T., Elliot, O., Patino-Galindo, J.A., Park, K., Rabadan, R., Honig, B., and Shapira, S.D. (2019). A structure-informed atlas of human-virus interactions. Cell *178*, 1526-1541. e1516.

Li, F. (2015). Receptor recognition mechanisms of coronaviruses: a decade of structural studies. Journal

of virology *89*, 1954-1964.

Masson, P., Hulo, C., De Castro, E., Bitter, H., Gruenbaum, L., Essioux, L., Bougueleret, L., Xenarios, I., and Le Mercier, P. (2012). ViralZone: recent updates to the virus knowledge resource. Nucleic acids research *41*, D579-D583.

Minor, P., Pipkin, P., Hockley, D., Schild, G., and Almond, J. (1984). Monoclonal antibodies which block cellular receptors of poliovirus. Virus research *1*, 203-212.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., and Dubourg, V. (2011). Scikit-learn: Machine learning in Python. Journal of machine learning research *12*, 2825-2830.

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

Qi, F., Qian, S., Zhang, S., and Zhang, Z. (2020). Single cell RNA sequencing of 13 human tissues identify cell types and receptors of human coronaviruses. bioRxiv.

Ryu, W.-S. (2016). Molecular virology of human pathogenic viruses (Academic Press).

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., and Tsafou, K.P. (2015). STRING v10: protein–protein interaction networks, integrated over the tree of life. Nucleic acids research *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.

Wu, A., Peng, Y., Huang, B., Ding, X., Wang, X., Niu, P., Meng, J., Zhu, Z., Zhang, Z., and Wang, J. (2020). Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. Cell Host & Microbe.

Yan, C., Duan, G., Wu, F.-X., and Wang, J. (2019). IILLS: predicting virus-receptor interactions based on similarity and semi-supervised learning. BMC bioinformatics *20*, 651.

Zhang, H., Kang, Z., Gong, H., Xu, D., Wang, J., Li, Z., Cui, X., Xiao, J., Meng, T., and Zhou, W. (2020). The digestive system is a potential route of 2019-nCov infection: a bioinformatics analysis based on single-cell transcriptomes. BioRxiv.

Zhang, Z., Zhu, Z., Chen, W., Cai, Z., Xu, B., Tan, Z., Wu, A., Ge, X., Guo, X., and Tan, Z. (2019). Cell membrane proteins with high n-glycosylation, high expression and multiple interaction partners are preferred by mammalian viruses as receptors. Bioinformatics *35*, 723-728.

Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., and Huang, C.-L. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature, 1-4.