

METHODOLOGY

Open Access



# IILLS: predicting virus-receptor interactions based on similarity and semi-supervised learning

Cheng Yan<sup>1,2</sup>, Guihua Duan<sup>1\*</sup>, Fang-Xiang Wu<sup>3</sup> and Jianxin Wang<sup>1</sup>

From Joint 30th International Conference on Genome Informatics (GIW) & Australian Bioinformatics and Computational Biology Society (ABACBS) Annual Conference  
Sydney, Australia. 9-11 December 2019

## Abstract

**Background:** Viral infectious diseases are the serious threat for human health. The receptor-binding is the first step for the viral infection of hosts. To more effectively treat human viral infectious diseases, the hidden virus-receptor interactions must be discovered. However, current computational methods for predicting virus-receptor interactions are limited.

**Result:** In this study, we propose a new computational method (IILLS) to predict virus-receptor interactions based on Initial Interaction scores method via the neighbors and the Laplacian regularized Least Square algorithm. IILLS integrates the known virus-receptor interactions and amino acid sequences of receptors. The similarity of viruses is calculated by the Gaussian Interaction Profile (GIP) kernel. On the other hand, we also compute the receptor GIP similarity and the receptor sequence similarity. Then the sequence similarity is used as the final similarity of receptors according to the prediction results. The 10-fold cross validation (10CV) and leave one out cross validation (LOOCV) are used to assess the prediction performance of our method. We also compare our method with other three competing methods (BRWH, LapRLS, CMF).

**Conclusion:** The experiment results show that IILLS achieves the AUC values of 0.8675 and 0.9061 with the 10-fold cross validation and leave-one-out cross validation (LOOCV), respectively, which illustrates that IILLS is superior to the competing methods. In addition, the case studies also further indicate that the IILLS method is effective for the virus-receptor interaction prediction.

**Keywords:** Virus-receptor interaction, Similarity, Semi-supervised learning, Laplacian regularized least squares classifier, Gaussian interaction profile (GIP) kernel

## Background

Viruses are the most abundant biological entities on the planet and widely distributed in organs of living organisms and environments [1, 2]. In particular, they are an important part of the human microbiome which is closely related with human health and diseases [3]. Actually, hundreds of human diseases were resulted from

viruses [4], such as Ebola virus (EBOV) [5], Zika virus [6], American Machupo virus (MACV), Guanarito virus (GTOV), Sabia virus (SABV), Junin virus (JUNV), and so on [7]. In marine environments, viruses can kill up to 40% of the standing stock of prokaryotes daily [8]. In addition, the cellular and physiological changes in the host cells can be caused by virus infections, such as altering genomic sequences and dysfunctioning their hosts [9, 10].

\*Correspondence: [duangh@csu.edu.cn](mailto:duangh@csu.edu.cn)

<sup>1</sup>School of Computer Science and Engineering, Central South University, 932 South Lushan Rd, 410083 ChangSha, China

Full list of author information is available at the end of the article



When viruses contact the surface of host cells, the virus process starts [11]. In general, the receptor-binding is considered as the first step for the viral infection of host cells [12]. The specificity and affinity are the main factors that viruses can use diverse types of molecules to attach to and enter into cells [13]. With the development of high-throughput technologies, many studies indicate that some molecules including proteins are the receptor of viruses [14], such as carbohydrates and lipids [15]. Furthermore, the virus-receptor interaction is also a dynamic process, as it can evolve over the course of an infection while virus variants with distinct receptor-binding specificity and tropism can appear [13]. In order to help understand the interaction mechanism between viruses and receptors, a database (called viralReceptor) with mammalian virus-receptor interactions has been constructed by Zhang et.al [16]. ViralReceptor consists of 128 viral species or sub-species, 119 receptors of mammalian and 268 interaction pairs between them. In addition, the structural and functional analysis of receptors also further provide the theoretic basis to discover new virus-receptor interactions, which include protein domains, higher level of N-glycosylation, higher ratio of self-interaction, and so on [16].

In this study, we propose a computational method (IILLS) based on Initial Interaction scores method via the neighbors and Laplacian regularized Least Square algorithm (a semi-supervised learning method), to predict virus-receptor interactions. IILLS integrates the known virus-receptor interactions and amino acid sequences of receptors to compute similarities of viruses and receptors. Then IILLS uses the Laplacian regularized Least Square algorithm and initial interaction scores based on the neighbors to construct the computational model. We conduct the 10-fold cross validation (10CV) and leave one out cross validation (LOOCV) to assess the prediction performance of IILLS and compare it with other three methods. The prediction performance of IILLS is best in terms of AUC (the area under of ROC curve) as its AUC values are 0.8675 and 0.9061 with 10CV and LOOCV, respectively. The evaluation results of case study also show that IILLS is an effective virus-receptor prediction method.

We also provide IILLS, via a web server, to predict virus-receptor interactions. The input of this web server is a receptor amino acid sequence or a txt file with multiple sequences in the FASTA format. The prediction result will be displayed after submission when uploading a sequence. However, the prediction results of the txt file of sequences is sent by the email with link page. Therefore, when uploading a sequence file, an email address should be provided. In addition, a job ID is assigned after one submission. According to job ID, the user can also obtain the prediction result from web server.

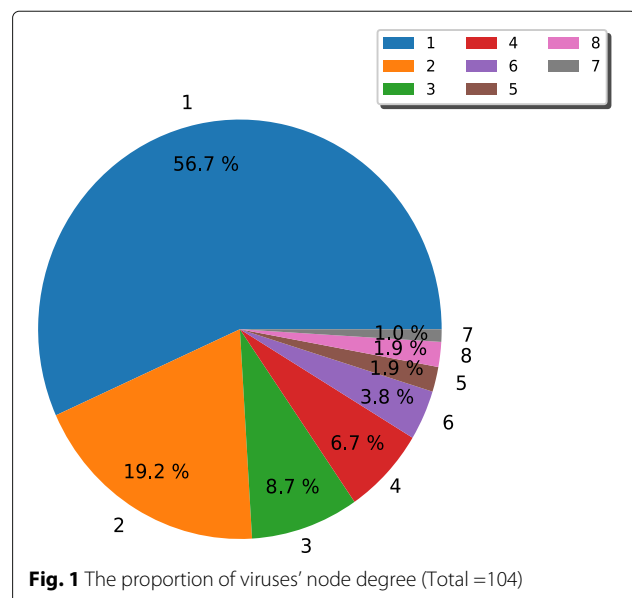
## Methods

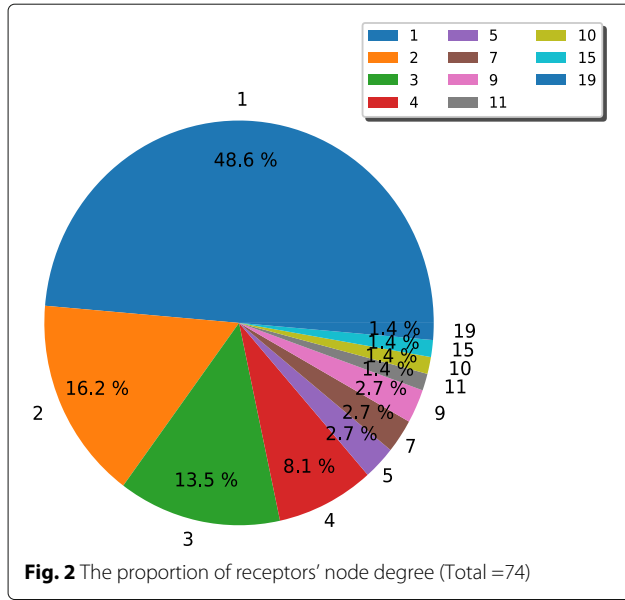
### Materials

We download the known mammalian virus-receptor interactions from viralReceptor database. Then we further extract human virus-receptor interactions as the benchmark dataset. It includes 104 virus species or sub-species, 74 receptors and 211 interaction pairs between viruses and receptors. The detail node degree distributions of viruses and receptors in this standard virus-receptor interaction network are also described in Figs. 1 and 2. The degree of a node is the number of edges which have this node as an endvertex in the virus-receptor interaction network. Each color represents the proportion of viruses (receptors) which have the same node degree. In Fig. 1, the node degrees of 104 virus range from 1 to 8, respectively. Their distribution proportion are 56.7%, 19.2%, 8.7%, 6.7%, 1.9%, 3.8%, 1.0% and 1.9%, respectively. In Fig. 2, each color represents the proportion of receptors with the same node degree. For example, the red color represents that 8.1% of all receptors have the node degree of 4.

### Similarity of viruses

Based on the assumption that similar viruses exhibit similar interaction profiles with receptors [17–20], we used the Gaussian Interaction Profile (GIP) similarity to measure the virus similarity. Let  $V = \{v_1, v_2, \dots, v_{N_v}\}$  be the set of  $N_v$  viruses,  $P = \{p_1, p_2, \dots, p_{N_p}\}$  be the set of  $N_p$  receptors, and  $Y \in R^{N_v \times N_p}$  be the adjacency matrix of the bipartite graph to describe known virus and receptor associations. When the virus  $v_i$  and receptor  $p_j$  have a known interaction, the value of  $y_{ij}$  is 1 and otherwise 0. The GIP similarity of viruses  $v_1$  and  $v_2$  can be computed as follows:





$$S_v(v_1, v_2) = G_v(v_1, v_2) = \exp(-\gamma_v \|y_{v_1} - y_{v_2}\|^2), \quad (1)$$

$$\gamma_v = \gamma'_v / \left( \frac{1}{N_v} \sum_{i=1}^{N_v} \|y_{v_i}\|^2 \right), \quad (2)$$

in which  $y_{v_1} = \{y_{11}, y_{12}, \dots, y_{1N_p}\}$  and  $y_{v_2} = \{y_{21}, y_{22}, \dots, y_{2N_p}\}$  are the interaction profiles of virus  $v_1$  and virus  $v_2$ , respectively. The parameter  $\gamma_v$  is used to regulate the kernel bandwidth. We can set the value of bandwidth parameter  $\gamma'_v$  by the cross validation. In this study, the parameter  $\gamma'_v$  is set to be 1 according to previous successful studies [17, 21, 22] and the influence analysis of prediction performance of parameter  $\gamma'_v$  by the 10-fold cross validation.

### Similarity of receptors

In this study, we take two methods to measure the receptor similarity, which include the GIP similarity and the amino acid sequence similarity. The GIP similarity of receptors is also computed by the known interactions of receptors. Specifically, for receptors  $p_1$  and  $p_2$ , their GIP similarity can be calculated as follows:

$$G_p(p_1, p_2) = \exp(-\gamma_p \|y_{p_1} - y_{p_2}\|^2), \quad (3)$$

$$\gamma_p = \gamma'_p / \left( \frac{1}{N_p} \sum_{i=1}^{N_p} \|y_{p_i}\|^2 \right), \quad (4)$$

in which  $y_{p_1} = \{y_{11}, y_{21}, \dots, y_{N_v,1}\}^T$  is the interaction profile of receptor  $p_1$  while  $y_{p_2} = \{y_{12}, y_{22}, \dots, y_{N_v,2}\}^T$  is the interaction profile of receptor  $p_2$ . Furthermore, the parameter  $\gamma_p$  is also used to control the kernel bandwidth and the parameter  $\gamma'_p$  is also set to be 1.

In addition, we compute the sequence similarity between receptors. First, we download the amino acid sequences of receptors from the KEGG GENE database [23]. The receptor sequence similarity is computed by their normalized Smith-Waterman score [24, 25]. For receptors  $p_1$  and  $p_2$ , the sequence similarity can be calculated as follows:

$$G_s(p_1, p_2) = SW(p_1, p_2) / \sqrt{SW(p_1, p_1)} \sqrt{SW(p_2, p_2)}, \quad (5)$$

in which  $SW(p_1, p_2)$  is the original Smith-Waterman score between receptor  $p_1$  and receptor  $p_2$ .

Based on the GIP similarity and the sequence similarity of receptors, we construct the final similarity of receptors  $S_p$  as follows:

$$S_p = \alpha * G_p + (1 - \alpha) * G_s, 0 \leq \alpha \leq 1.0 \quad (6)$$

where  $\alpha$  is the weight parameter.

### Initialized interaction profiles for new viruses and receptors

The quality of known virus-receptors has important impact on the performance of prediction method. In this study, we want to set the initialized interaction scores for viruses (receptors) which have no known interaction with receptors (viruses). Inspired by the KNN method, we take the interaction profiles of all neighbors into consideration, which have known interactions. For example, the initial interaction profile between a new virus  $v_i$  and receptor  $p_j$  can be calculated as follows:

$$y(v_i, p_j) = \frac{\sum_{l=1}^{N_v} S_v^{(il)} y_{lj}}{\sum_{l=1}^{N_v} S_v^{(il)}} \quad (7)$$

in which  $S_v^{(il)}$  is the GIP similarity between viruses  $v_i$  and  $v_l$ .

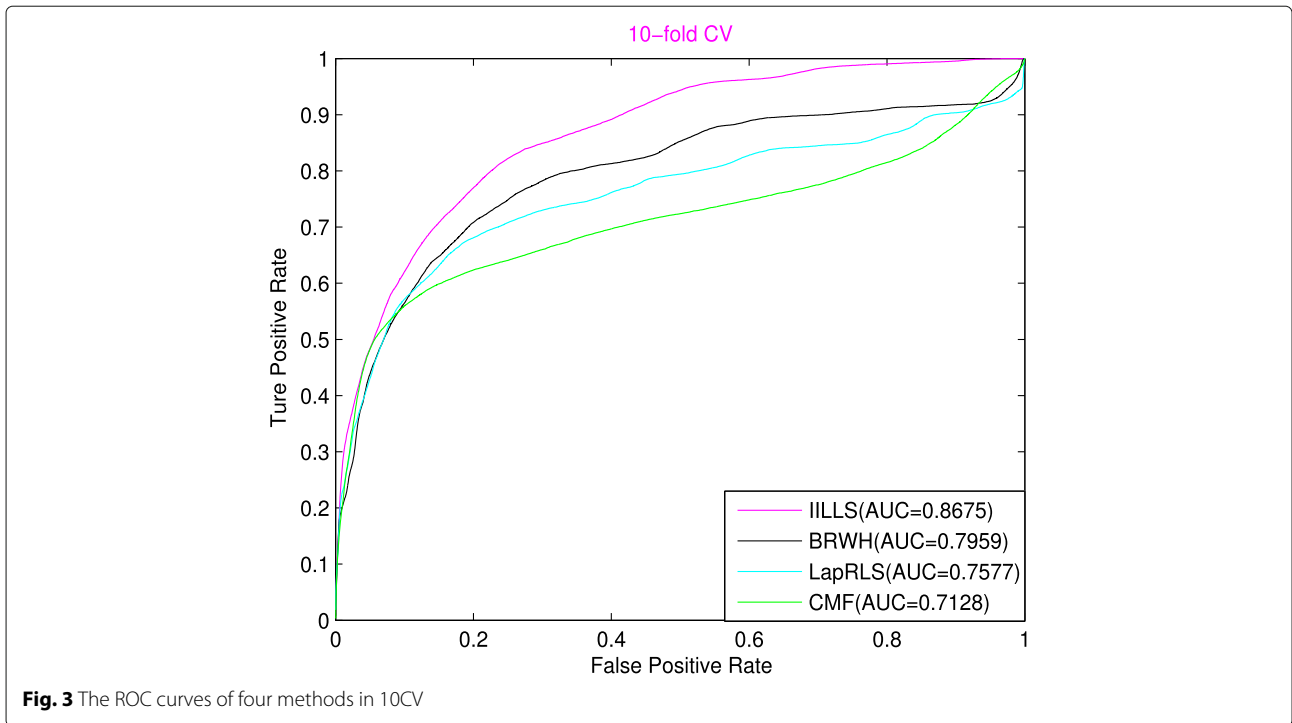
Similarly, we also apply the same model to calculate the interaction profiles of new receptor. Specifically, the initial interaction profile between virus  $v_i$  and a new receptor  $p_j$  can be calculated as follows:

$$y(v_i, p_j) = \frac{\sum_{l=1}^{N_p} S_p^{(jl)} y_{il}}{\sum_{l=1}^{N_p} S_p^{(jl)}} \quad (8)$$

in which  $S_p^{(jl)}$  is the final similarity between receptors  $p_j$  and  $p_l$ .

### Laplacian regularized least square for virus-receptor interaction prediction

Inspired by successful applications of Laplacian regularized Least Square (LapRLS) model in predicting drug-target interactions [26–28], we adopt the LapRLS model



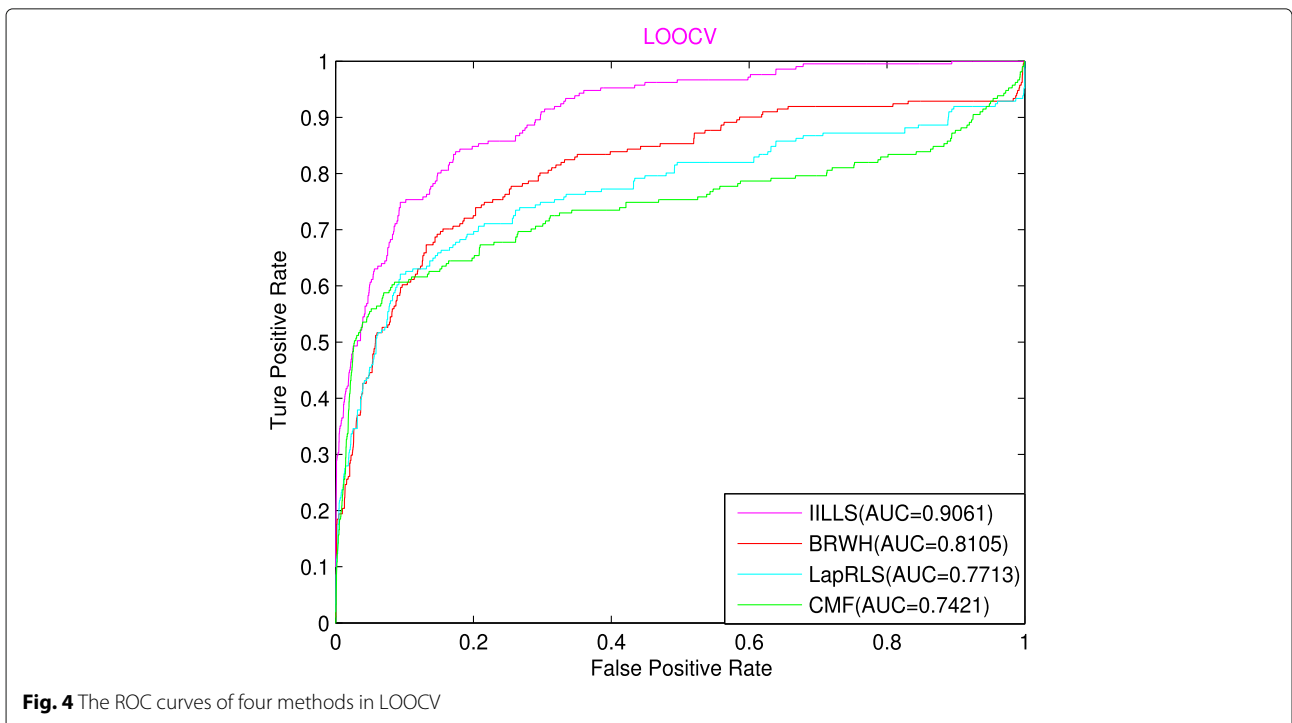
**Fig. 3** The ROC curves of four methods in 10CV

to predict virus-receptor interactions. After obtaining the similarity matrices, we construct the normalized Laplacian matrices for viruses and receptors as follows:

$$L^v = (D^v)^{-1/2}(D^v - S_v)(D^v)^{-1/2}, \tag{9}$$

$$L^p = (D^p)^{-1/2}(D^p - S_p)(D^p)^{-1/2}, \tag{10}$$

where the matrix  $D^v$  is the diagonal matrix whose element  $D^v(i, i)$  is calculated by the sum of row  $i$  of the virus similarity matrix  $S_v$ . Similarly, the matrix



**Fig. 4** The ROC curves of four methods in LOOCV

$D^p$  is calculated based on the receptor similarity matrix  $S_p$ .

For viruses and receptors, prediction matrixes  $F_v$  and  $F_p$  are respectively calculated from the LapRLS model by minimizing the cost functions as follows:

$$F_v^* = \arg \min_{F_v} \left[ \|Y - F_v\|_F^2 + \beta_v \text{tr} \left( F_v^T L^v F_v \right) \right], \quad (11)$$

$$F_p^* = \arg \min_{F_p} \left[ \|Y - F_p\|_F^2 + \beta_p \text{tr} \left( F_p^T L^p F_p \right) \right], \quad (12)$$

in which  $\text{tr}(\cdot)$  is the trace of a matrix,  $Y$  is the adjacency matrix of the known virus-receptor interactions,  $L_v$  and  $L_p$  are the normalized Laplacian matrices of virus similarity and receptor similarity, and  $\|\cdot\|_F$  is the Frobenius norm.  $\beta_v$  and  $\beta_p$  are the trade-off parameters and are set to be 1. According to previous studies [29], the computation model can be solved by:

$$F_v^* = S^v (S^v + \beta_v L^v S^v)^{-1} Y, \quad (13)$$

$$F_p^* = S^p (S^p + \beta_p L^p S^p)^{-1} Y^T, \quad (14)$$

Finally, we obtain the virus-receptor interaction prediction matrix  $F^*$  by the mean of results of viruses and receptors:

$$F^* = \left( F_v^* + (F_p^*)^T \right) / 2. \quad (15)$$

## Results

### Performance evaluation

In order to assess the prediction performance of IILLS, we conduct the 10CV and LOOCV. The AUC is the metric to evaluate the prediction performance. We compare our method with other three methods: BRWH [30], LapRLS [26] and CMF [31].

### Comparison with other methods

Figure 3 shows the prediction performance of four methods in 10CV. Compared with other methods (BRWH: 0.7959, LapRLS: 0.7577, CMF: 0.7128), IILLS achieves the best prediction performance with the AUC value of 0.8675.

Figure 4 also shows that IILLS is superior to other methods in terms of AUC values (IILLS: 0.9061, BRWH:

0.8105, LapRLS: 0.7713, CMF: 0.7421). These experiment results illustrate that IILLS can obtain the better prediction performance.

### Analyzing receptor similarity

In this study, we also analyze the receptor similarity based on the GIP similarity and sequence similarity in terms of the influences of prediction performance of parameter  $\alpha$  in our method. We conduct 10CV and LOOCV to compute the prediction performance.

Table 1 shows the 10CV prediction performances of various parameter values of  $\alpha$  ranging from 0 to 1.0 with the increment of 0.1. We can see from Table 1 that our method obtains the best prediction performance in 10CV when only using sequence similarity ( $\alpha = 0$ ). The AUC value of our method has a slightly descending trend when  $\alpha$  ranges from 0 to 1.0.

Table 2 shows the LOOCV prediction performances of various parameter values of  $\alpha$  ranging from 0 to 1.0 with the increment of 0.1. We can see from Table 2 that our method also obtains the best prediction performance in LOOCV when only using sequence similarity ( $\alpha = 0$ ). The AUC value of our method has also a slightly descending trend when  $\alpha$  ranges from 0 to 1.0. Therefore, we set the  $\alpha$  to be 0 in this study.

In addition, we also provide the ROC of our method on different values of parameter  $\alpha$  in three cases. The first only uses the sequence similarity of receptors ( $\alpha = 0$ ). The second only uses the GIP similarity of receptors ( $\alpha = 1.0$ ). The third is with the mean of GIP similarity and sequence similarity of receptors ( $\alpha = 0.5$ ).

Figures 5 and 6 show the prediction performances of IILLS under three different receptor similarities in 10CV and LOOCV, respectively. We can also see from Figs. 5 and 6 that IILLS achieves the best prediction performance when only using the sequence similarity.

### Parameter analysis for $\gamma'_v$

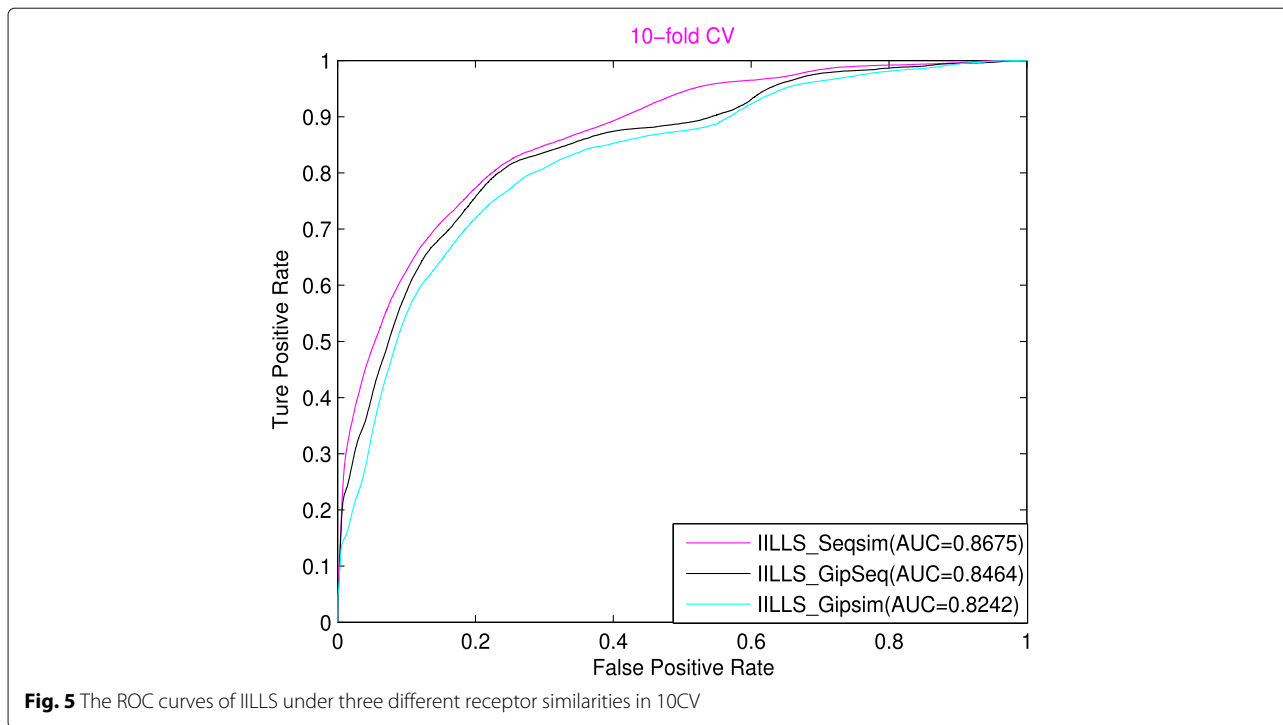
In this section, we analyze parameters  $\gamma'_v$ . In addition, by considering the effect of parameter  $\gamma'_v$  is similar to the effect of parameter  $\gamma'_p$ , we set  $\gamma'_p = \gamma'_v$ . When only using the sequence similarity, Table 3 shows the 10CV prediction performances of value set (0.25, 0.5, 1, 2, 4) of parameter  $\gamma'_v$ . We can see from Table 3 that our method

**Table 1** The 10CV prediction performances of various parameter values of  $\alpha$  ranging from 0 to 1.0 with the increment of 0.1, the best result is in the bold face

$\alpha$	0	0.1	0.2	0.3	0.4
AUC	<b>0.8675</b>	0.8611	0.8544	0.8500	0.8475
0.5	0.6	0.7	0.8	0.9	1.0
0.8464	0.8425	0.8417	0.8376	0.8327	0.8242

**Table 2** The LOOCV prediction performances of various parameter values of  $\alpha$  ranging from 0 to 1.0 with the increment of 0.1, the best result is in the bold face

$\alpha$	0	0.1	0.2	0.3	0.4
AUC	<b>0.9061</b>	0.8975	0.8935	0.8905	0.8885
0.5	0.6	0.7	0.8	0.9	1.0
0.8865	0.8846	0.8828	0.8806	0.8779	0.8724

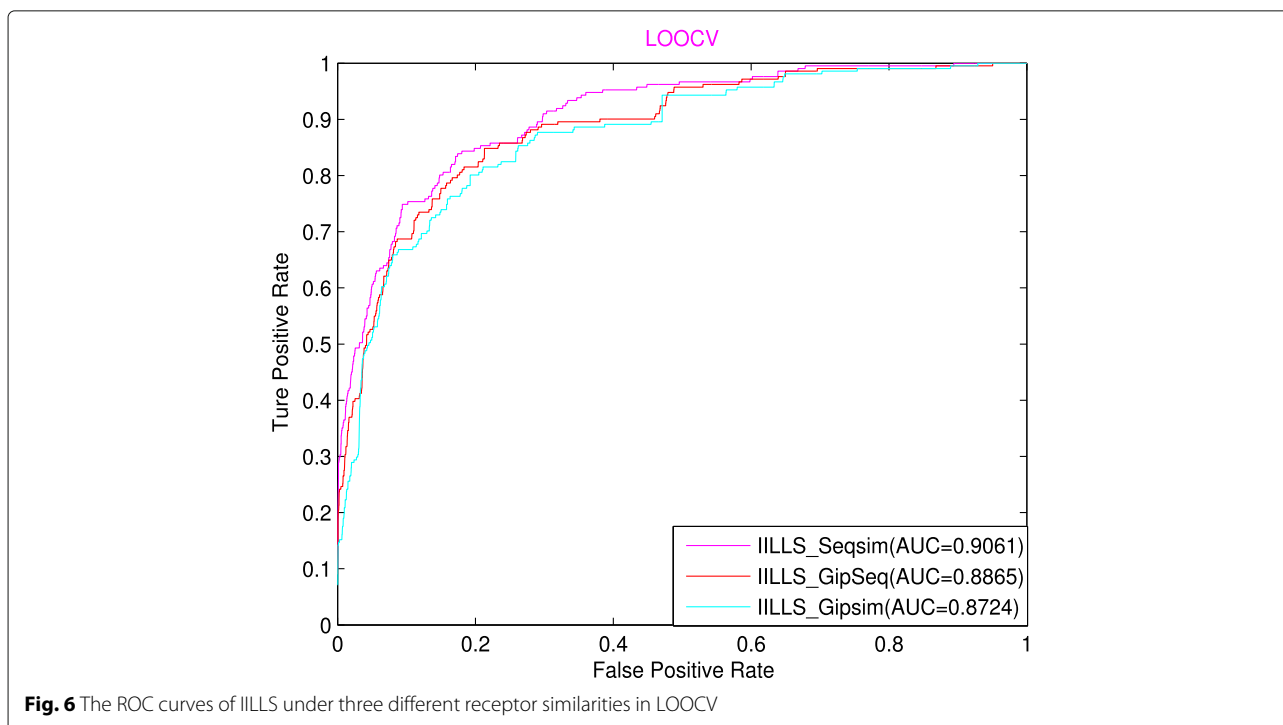


**Fig. 5** The ROC curves of ILLS under three different receptor similarities in 10CV

obtains best prediction performance in 10CV when  $\gamma'_v$  is set to be 2. The AUC value under setting  $\gamma'_v = 2$  is slightly better than the AUC value when  $\gamma'_v = 1$ . Therefore, we also simply set the  $\gamma'_v = 1$  as the default value based on the previous successful studies and experiment results of 10CV.

**Case studies**

In order to further evaluate the prediction performance of ILLS in applications, we analyze the prediction ability of our method in discovering new virus-receptor interactions. The extracted human virus-receptor interactions are used as the benchmark datasets.



**Fig. 6** The ROC curves of ILLS under three different receptor similarities in LOOCV

**Table 3** The 10CV prediction performances of various parameter values of  $\gamma_v$ , the best result is in the bold face

$\gamma_v$	0.25	0.5	1	2	4
AUC	0.8550	0.8608	0.8675	<b>0.8700</b>	0.8434

Table 4 shows the validation results of top 10 virus-receptor interactions which are predicted by IILLS. We can see from Table 4 that 5 of 10 predicted associations are validated by previous studies. C-type lectin domain family 4 member M (CLEC4M, also called L-SIGN or CD209L) is equipped with a carbohydrate recognition domain (CRD) that mediates the recognition of fucose and high-mannose glycans in a Ca<sup>2+</sup>-dependent manner, these carbohydrate structures can be found in multiple pathogens, such as Lassa virus, Ebola virus, among others [32, 33]. The CD209 is also the receptor of known SARS-CoV, human coronaviruses and 229E, although the disease caused by SARS-CoV differs from the diseases caused by the known human coronaviruses and 229E [34]. L-SIGN (also called DC-SIGN) is related to CLEC4M and is a C-type lectin involved in both innate and adaptive immunity, they are known to bind to multiple pathogens and function as cellular receptors for various viruses, such as Dengue virus [35]. Rift Valley fever virus (RVFV) goes through L-SIGN to infect cells expressing the lectin ectopically [32, 36]. The phleboviruses, such as Uukuniemi virus (UUKV), can exploit L-SIGN for infection [32, 36].

## Discussion

With the development of high-through sequencing technology and microbiology, many studies have evidenced that microbes have key impacts on health body and human diseases. Furthermore, the viruses are an important part of the human microbiomes, and are also the

direct origin of infectious diseases, such as Sabia virus and so on. The receptor-binding is the first step for viral infection of host cells. Therefore, in order to systematically understand the mechanisms between virus and receptor and improve the diagnosis and treatment of infectious diseases, it need develop effective methods to identify new virus-receptor interactions.

## Conclusion

In this study, we develop a computational method (IILLS) to predict virus-receptor interactions of human with known virus-receptor interactions and the amino acid sequence of receptors. Firstly, IILLS computes the virus similarity by GIP kernel. Then we also calculate the receptor GIP kernel similarity and the receptor sequence similarity. The final receptor similarity is constructed by the sequence similarity based on the experiment results. IILLS uses the Laplacian regularized Least Square (LapRLS) model to predict the potential virus-disease interactions. It further improves the prediction performance by adding an initial interaction scores process for new viruses and receptors. In terms of AUC with 10CV and LOOCV, IILLS can achieves better prediction performance than other three competing methods. The case studies also show that IILLS can effectively predict virus-receptor interactions, and also help control the virus infectious diseases in the future.

However, there still exist some limitations in IILLS. On the one hand, the virus similarity is calculated by the GIP kernel with known virus-receptor interactions. We should consider more relevant biological network information, such as sequence information. In addition, other integration methods of receptor similarity also should be considered in the future. Finally, other latest matrix factorization methods also should be considered, such

**Table 4** The validated result of top 10 predicted virus-receptor interactions

Rank	Virus	Receptor	References
1	Lymphocytic choriomeningitis mammarenavirus (LCMV)	C-type lectin domain family 4 member M(CLEC4M, L-SIGN)	Unknown
2	Lassa mammarenavirus	C-type lectin domain family 4 member M	Garcia-Vallejo et al, (2015) and Sakuntabhai et al, (2005)
3	Human coronavirus 229E (229E)	CD209 molecule (CD209)	Lo et al., (2006)
4	Dengue virus	C-type lectin domain family 4 member M	Li et al., (2012)
5	Rift Valley fever virus	C-type lectin domain family 4 member M	Lger et al., (2016), and Sakuntabhai et al., (2005)
6	Uukuniemi virus	C-type lectin domain family 4 member M	Lger et al., (2016), and Sakuntabhai et al., (2005)
7	Human immunodeficiency virus 2	C-type lectin domain family 4 member M	Unknown
8	Human alphaherpesvirus 1	integrin subunit beta 3 (beta 3 integrin)	Unknown
9	Coxsackievirus A9 (CAV9)	integrin subunit beta 1	Unknown
10	Human betaherpesvirus 5	integrin subunit beta 6	Unknown

as DNRLMF-MDA [37], DRRS [38], SIMCLDA[39] and BNNR [40]. Therefore, we would like to develop a more effective method for predicting virus-receptor interactions by addressing the above limitations in the future.

#### Abbreviations

10CV: 10-fold cross validation; AUC: Area under the receiver operating curve; BRWH: Bi-random walk on a heterogeneous network; CMF : Collaborative matrix factorization; GIP: Gaussian interaction profile; KNN: K-nearest neighbors; LapRLS: Laplacian regularized least squares classifier; SW: Normalized Smith-Waterman score

#### Acknowledgements

The authors are very grateful to the anonymous reviewers for their constructive comments which have helped significantly in revising this work. The authors would like to express their gratitude for the support from the National Natural Science Foundation of China( No.61772552, No.61420106009, No.61832019 and No.61962050), 111 Project (No. B18059) and Hunan Provincial Science and Technology Program (No. 2018WK4001).

#### About this supplement

This article has been published as part of *BMC Bioinformatics Volume 20 Supplement 23, 2019: Proceedings of the Joint International GIW & ABACBS-2019 Conference: bioinformatics*. The full contents of the supplement are available online at <https://bmcbioinformatics.biomedcentral.com/articles/supplements/volume-20-supplement-23>.

#### Authors' contributions

JW conceived the project; CY designed the experiments; and CY performed the experiments; CY, GD and FXW wrote the paper. All authors read and approved the final manuscript.

#### Funding

Publication costs are funded by National Natural Science Foundation of China under Grant No.61420106009.

#### Availability of data and materials

The web server of ILLS method is available at <http://bioinformatics.csu.edu.cn/ILLIS>.

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>School of Computer Science and Engineering, Central South University, 932 South Lushan Rd, 410083 ChangSha, China. <sup>2</sup>School of Computer and Information, Qiannan Normal University for Nationalities, Longshan Road, 558000 DuYun, China. <sup>3</sup>Biomedical Engineering and Department of Mechanical Engineering, University of Saskatchewan, SKS7N5A9, Saskatoon Canada.

Received: 14 November 2019 Accepted: 18 November 2019

Published: 27 December 2019

#### References

- Minot S, Sinha R, Chen J, Li H, Keilbaugh SA, Wu GD, Lewis JD, Bushman FD. The human gut virome: inter-individual variation and dynamic response to diet. *Genome Res.* 2011;21(10):1616–25.
- Paez-Espino D, Eloë-Fadros EA, Pavlopoulos GA, Thomas AD, Huntemann M, Mikhailova N, Rubin E, Ivanova NN, Kyrpides NC. Uncovering earth's virome. *Nature.* 2016;536(7617):425.
- Wigington CH, Sonderegger D, Brussaard CP, Buchan A, Finke JF, Fuhrman JA, Lennon JT, Middelboe M, Suttle CA, Stock C, et al. Re-examination of the relationship between marine virus and microbial cell abundances. *Nat Microbiol.* 2016;1(3):15024.
- Geoghegan JL, Senior AM, Di Giallonardo F, Holmes EC. Virological factors that increase the transmissibility of emerging human viruses. *Proc Natl Acad Sci.* 2016;113(15):4170–5.
- Maganga GD, Kapetshi J, Berthet N, Kebela Ilunga B, Kabange F, Mbala Kingebeni P, Mondonge V, Muyembe J-JT, Bertherat E, Briand S, et al. Ebola virus disease in the democratic republic of congo. *New England J Med.* 2014;371(22):2083–91.
- Mlakar J, Korva M, Tul N, Popović M, Poljšak-Prijatelj M, Mraz J, Kolenc M, Resman Rus K, Vesnaver Vipotnik T, Fabjan Vodušek V, et al. Zika virus associated with microcephaly. *New England J Med.* 2016;374(10):951–8.
- Moraz M-L, Kunz S. Pathogenesis of arenavirus hemorrhagic fevers. *Expert Rev Anti-Infect Ther.* 2011;9(1):49–59.
- Suttle CA. Marine viruses—major players in the global ecosystem. *Nat Rev Microbiol.* 2007;5(10):801.
- Qin N, Yang F, Li A, Prifti E, Chen Y, Shao L, Guo J, Le Chatelier E, Yao J, Wu L, et al. Alterations of the human gut microbiome in liver cirrhosis. *Nature.* 2014;513(7516):59.
- Cadwell K. The virome in host health and disease. *Immunity.* 2015;42(5):805–13.
- Boulant S, Stanifer M, Lozach P-Y. Dynamics of virus-receptor interactions in virus binding, signaling, and endocytosis. *Viruses.* 2015;7(6):2794–815.
- Baranowski E, Ruiz-Jarabo CM, Domingo E. Evolution of cell recognition by viruses. *Science.* 2001;292(5519):1102–5.
- Casasnovas JM. Virus-receptor interactions and receptor-mediated virus entry into host cells. *Subcell Biochem.* 2013;68:441–66.
- Li F. Structure, function, and evolution of coronavirus spike proteins. *Ann Rev Virol.* 2016;3:237–61.
- Peng W, de Vries RP, Grant OC, Thompson AJ, McBride R, Tsogtbaatar B, Lee PS, Razi N, Wilson IA, Woods RJ, et al. Recent h3n2 viruses have evolved specificity for extended, branched human-type receptors, conferring potential for increased avidity. *Cell Host Microbe.* 2017;21(1):23–34.
- Zhang Z, Zhu Z, Chen W, Cai Z, Xu B, Tan Z, Wu A, Ge X, Guo X, Tan Z, et al. Cell membrane proteins with high n-glycosylation, high expression and multiple interaction partners are preferred by mammalian viruses as receptors. *Bioinformatics.* 2018;35(5):723–8.
- Laarhoven TV, Nabuurs SB, Marchiori E. Gaussian interaction profile kernels for predicting drug-target interaction. *Bioinformatics.* 2011;27(21):3036–43.
- Yan C, Guihua D, Wu FX, Pan Y, Wang J. Brwmda: predicting microbe-disease associations based on similarities and bi-random walk on disease and microbe networks. *IEEE/ACM Trans Comput Biol Bioinform.* 2019. <https://doi.org/10.1109/TCBB.2019.2907626>.
- Yan C, Wang J, Wu F-X. Dwnn-rls: regularized least squares method for predicting circrna-disease associations. *BMC Bioinformatics.* 2018;19(19):520.
- Yan C, Duan G, Wu F, Pan Y, Wang J. Mchmda: Predicting microbe-disease associations based on similarities and low-rank matrix completion. *IEEE/ACM Trans Comput Biol Bioinform.* 2019. <https://doi.org/10.1109/TCBB.2019.2926716>.
- Lan W, Wang J, Li M, Liu J, Wu F-X, Pan Y. Predicting microRNA-disease associations based on improved microRNA and disease similarities. *IEEE/ACM Trans Comput Biol Bioinform.* 2018;15(6):1774–82.
- Lan W, Li M, Zhao K, Liu J, Wu F-X, Pan Y, Wang J. Ldap: a web server for lncrna-disease association prediction. *Bioinformatics.* 2016;33(3):458–60.
- Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M. From genomics to chemical genomics: new developments in kegg. *Nucleic Acids Res.* 2006;34(suppl\_1):354–7.
- Smith TF, Waterman MS, et al. Identification of common molecular subsequences. *J Mol Biol.* 1981;147(1):195–7.
- Jiang H, Wang J, Li M, Lan W, Wu F, Pan Y. mirts: A recommendation algorithm for predicting miRNA targets. *IEEE/ACM Trans Comput Biol Bioinform.* 2018. <https://doi.org/10.1109/TCBB.2018.2873299>.
- Xia Z, Wu L-Y, Zhou X, Wong ST. Semi-supervised drug-protein interaction prediction from heterogeneous biological spaces. *BMC Syst Biol.* 2010;4:6. BioMed Central.



27. Yuan Q, Gao J, Wu D, Zhang S, Mamitsuka H, Zhu S. Druge-rank: improving drug–target interaction prediction of new candidate drugs or targets by ensemble learning to rank. *Bioinformatics*. 2016;32(12):18–27.
28. Yan C, Wang J, Lan W, Wu F-X, Pan Y. Sdtrls: Predicting drug-target interactions for complex diseases based on chemical substructures. *Complexity*. 2017;2017(Article ID 2713280):10.
29. Belkin M, Niyogi P, Sindhwani V. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *J Mach Learn Res*. 2006;7(Nov):2399–434.
30. Luo H, Wang J, Li M, Luo J, Peng X, Wu F-X, Pan Y. Drug repositioning based on comprehensive similarity measures and bi-random walk algorithm. *Bioinformatics*. 2016;32(17):2664–71.
31. Zheng X, Ding H, Mamitsuka H, Zhu S. Collaborative matrix factorization with multiple similarities for predicting drug-target interactions. In: *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM; 2013. p. 1025–33. <https://doi.org/10.1145/2487575.2487670>.
32. Sakuntabhai A, Turbpaiboon C, Casadémont I, Chuansumrit A, Lowhnoo T, Kajaste-Rudnitski A, Kalayanarooj SM, Tangnararatchakit K, Tangthawornchaikul N, Vasanawathana S, et al. A variant in the cd209 promoter is associated with severity of dengue disease. *Nat Genet*. 2005;37(5):507.
33. Garcia-Vallejo JJ, van Kooyk Y. Dc-sign: the strange case of dr. jekyll and mr. hyde. *Immunity*. 2015;42(6):983–5.
34. Lo AW, Tang NL, To K-F. How the sars coronavirus causes disease: host or organism?. *J Pathol J Pathol Soc Great B Irel*. 2006;208(2):142–51.
35. Li H, Wang J-X, Wu D-D, Wang H-W, Tang NL-S, Zhang Y-P. The origin and evolution of variable number tandem repeat of clec4m gene in the global human population. *PLoS ONE*. 2012;7(1):30268.
36. Léger P, Tetard M, Youness B, Cordes N, Rouxel RN, Flamand M, Lozach P-Y. Differential use of the c-type lectins I-sign and dc-sign for phlebovirus endocytosis. *Traffic*. 2016;17(6):639–56.
37. Yan C, Wang J, Ni P, Lan W, Wu F, Pan Y. Dnrlmf-mda: Predicting microrna-disease associations based on similarities of micrnas and diseases. *IEEE/ACM Trans Comput Biol Bioinform*. 2019;16(1):233–43.
38. Luo H, Li M, Wang S, Liu Q, Li Y, Wang J. Computational drug repositioning using low-rank matrix approximation and randomized algorithms. *Bioinformatics*. 2018;34(11):1904–12.
39. Lu C, Yang M, Luo F, Wu F-X, Li M, Pan Y, Li Y, Wang J. Prediction of Incrna–disease associations based on inductive matrix completion. *Bioinformatics*. 2018;34(19):3357–64.
40. Yang M, Luo H, Li Y, Wang J. Drug repositioning based on bounded nuclear norm regularization. *Bioinformatics*. 2019;35(14):455–63.

## Publisher's Note

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

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

