frontiers | Frontiers in Microbiology

TYPE Methods
PUBLISHED 05 December 2022
DOI 10.3389/fmicb.2022.1062281

OPEN ACCESS

EDITED BY
Fei Ma,
Chinese Academy of Medical Sciences and Peking Union Medical College, China

REVIEWED BY
Chirasmita Nayak,
Alagappa University,
India
Guohua Huang,
Shaoyang University,
China

*CORRESPONDENCE
Ju Xiang
xiang.ju@foxmail.com
Jianjun He
hejianjun@csmu.edu.cn
Binsheng He
hbscsmu@163.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Systems Microbiology, a section of the journal Frontiers in Microbiology

RECEIVED 05 October 2022
ACCEPTED 21 November 2022
PUBLISHED 05 December 2022

CITATION
Wang Y, Xiang J, Liu C, Tang M, Hou R, Bao M, Tian G, He J and He B (2022) Drug repositioning for SARS-CoV-2 by Gaussian kernel similarity bilinear matrix factorization.
Front. Microbiol. 13:1062281.
doi: 10.3389/fmicb.2022.1062281

Drug repositioning for SARS-CoV-2 by Gaussian kernel similarity bilinear matrix factorization

Yibai Wang[1†], Ju Xiang[1,2*†], Cuicui Liu[1], Min Tang[3], Rui Hou[4,5], Meihua Bao[6,7], Geng Tian[4,5], Jianjun He[2,6,7*] and Binsheng He[2,6,7*]

[1]School of Information Engineering, Changsha Medical University, Changsha, China, [2]Academician Workstation, Changsha Medical University, Changsha, China, [3]School of Life Sciences, Jiangsu University, Zhenjiang, Jiangsu, China, [4]Geneis (Beijing) Co., Ltd., Beijing, China, [5]Qingdao Geneis Institute of Big Data Mining and Precision Medicine, Qingdao, China, [6]School of Pharmacy, Changsha Medical University, Changsha, China, [7]Key Laboratory Breeding Base of Hunan Oriented Fundamental and Applied Research of Innovative Pharmaceutics, Changsha Medical University, Changsha, China

Coronavirus disease 2019 (COVID-19), a disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is currently spreading rapidly around the world. Since SARS-CoV-2 seriously threatens human life and health as well as the development of the world economy, it is very urgent to identify effective drugs against this virus. However, traditional methods to develop new drugs are costly and time-consuming, which makes drug repositioning a promising exploration direction for this purpose. In this study, we collected known antiviral drugs to form five virus-drug association datasets, and then explored drug repositioning for SARS-CoV-2 by Gaussian kernel similarity bilinear matrix factorization (VDA-GKSBMF). By the 5-fold cross-validation, we found that VDA-GKSBMF has an area under curve (AUC) value of 0.8851, 0.8594, 0.8807, 0.8824, and 0.8804, respectively, on the five datasets, which are higher than those of other state-of-art algorithms in four datasets. Based on known virus-drug association data, we used VDA-GKSBMF to prioritize the top-k candidate antiviral drugs that are most likely to be effective against SARS-CoV-2. We confirmed that the top-10 drugs can be molecularly docked with virus spikes protein/human ACE2 by AutoDock on five datasets. Among them, four antiviral drugs ribavirin, remdesivir, oseltamivir, and zidovudine have been under clinical trials or supported in recent literatures. The results suggest that VDA-GKSBMF is an effective algorithm for identifying potential antiviral drugs against SARS-CoV-2.

KEYWORDS

SARS-CoV-2, drug repositioning, bilinear matrix factorization, molecular docking, machine learning

Introduction

Caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a new infectious disease called coronavirus disease 2019 (COVID-19) has caused a big pandemic worldwide since 2019 (Eurosurveillance editorial team, 2020; Cheng et al., 2021a; Zhang et al., 2021). SARS-CoV-2 can transmit by human-to-human contacts, and is currently spreading rapidly to more than 400 countries around the world, causing millions of deaths (Coronaviridae Study Group of the International Committee on Taxonomy of V, 2020; Li et al., 2020; Cohain et al., 2021). Thus, SARS-CoV-2 seriously threatens human life and health as well as the development of world economy (Wu et al., 2020; Zhou P. et al., 2020; Zhu et al., 2020; Cheng et al., 2021b), and it is critical to find effective measures to prevent the transmission and fight against this virus.

One effective way to prevent the transmission of a virus is through vaccination. However, viruses like SARS-CoV-2 and influenzas are under rapid genetic and antigenic evolution, especially in their spike proteins (Yao et al., 2017; Zhang et al., 2017), which will make the vaccine less effective. Another method is to develop specific drug against the viruses. However, traditional methods to develop new drugs usually take years and cost tens of millions of dollars (Novac, 2013). With the development of various computational algorithms for mining intrinsic associations in biomedical data (Zhang et al., 2019; Xu et al., 2020a; Liu et al., 2021; Xiang et al., 2021a, 2022b; He et al., 2022; Yang et al., 2022), drug repositioning has become an effective way of exploring new uses for approved drugs, since it can significantly reduce the time and cost in the development of drugs (Liu et al., 2016, 2020; Yang J. et al., 2020; Zhu et al., 2021).

There are a few studies to prioritize approved drugs against SARS-CoV-2. For example, Zhou et al. proposed a KATZ method to probe antiviral drugs against SARS-CoV-2 through virus-drug association prediction (Zhou L. et al., 2020). More recently, Tang et al. prioritized drugs for COVID-19 through an indicator regularized non-negative matrix factorization method (Tang et al., 2020). Peng et al. collected an antivirial drug database and minied it to repurpose drugs aginst SARS-CoV-2 (Peng et al., 2020; Zhou L. et al., 2020). Wang et al. predicted anti-SARS-COV-2 drugs by bound nuclear norm regularization (Wang et al., 2021). Meng et al. builded the human drug virus database and identified anti-SARS-COV-2 drugs by similarity constrained probabilistic matrix factorization (Lu et al., 2021; Meng et al., 2021; Parsza et al., 2021). Shen et al. prioritized anti-SARS-CoV-2 drugs by combining an unbalanced bi-random walk and Laplacian regularized least squares (Shen et al., 2022). Though these methods achieved relatively good prediction performance in cross-validation and literature mining, the accuracy of prediction is yet to be improved and a more robust validation method is needed for further wet-lab experiments. Therefore, in this study, we collected the data of well-studied viruses that are similar to SARS-CoV-2 and their known antiviral drugs, forming a virus-drug association matrix (VDA). Then, we proposed a novel method for exploring potential virus-drug associations of SARS-CoV-2 by using Gaussian kernel similarity bilinear matrix factorization (VDA-GKSBMF).

The rest of the work is organized as follows. First, we collect five datasets and propose the details of the VDA-GKSBMF method for predicting potential virus-drug associations of SARS-CoV-2. Then, we study the effectiveness of the method by the 5-fold cross-validation experiments and compare VDA-GKSBMF with other state-of-art algorithms. Based on known virus-drug association data, we use VDA-GKSBMF to prioritize top-10 candidate antiviral drugs that are most likely to fight against SARS-CoV-2, and then evaluate the molecular binding activity between predicted antiviral drugs and SARS-CoV-2 spike protein (Gralinski, 2020) or human ACE2 (Zhao et al., 2020), to confirm whether the top-10 drugs are to be molecularly docked with the virus spikes protein or human ACE2. We also explore literatures to check if the top predicted drugs are under clinical trials or experiments against SARS-CoV-2.

Materials and methods

The overall workflow of the method is illustrated in Figure 1. We first introduce the datasets in this study, and then describe the details of the VDA-GKSBMF method for drug repositioning of SARS-CoV-2, including the construction of virus–drug heterogeneous network and the VDA-GKSBMF model, along with the alternating direction method of multipliers (ADMM) for solving the model to fill out unknown associations in virus–drug matrix.

Materials

To identify potential VDAs involving SARS-COV-2, we collect five datasets. There is Virus similarity matrix, drug similarity matrix, and VDA matrix in each dataset. Viruses are similar to SARS-CoV-2, small-molecule drugs and VDAs between them from the DrugBank (Wishart et al., 2018), PubChem (Kim et al., 2016), and NCBI (Wheeler et al., 2004) databases (see Table 1 for details).

These VDAs are represented by a VDA matrix $B_{m\times n}$, where $B_{dv}=1$ if the d-th drug is associated with the v-th virus, otherwise, $B_{dv}=0$. This forms a virus-drug association network, which can be denoted as a bipartite graph $G(V,D,E)$, where $E(G)=\{e_{ij}\}\subseteq V\times D$ contains edges representing known associations between viruses and drugs.

For viruses, we obtain the sequence-based similarities between viruses that are calculated by MAFFT (Katoh and Toh, 2008). For drugs, we obtain the chemical structure-based similarity scores between drugs by RDKit (Landrum, 2014), where chemical structures of drugs are obtained from the DrugBank database (Wishart et al., 2018). The details are shown in Table 1.

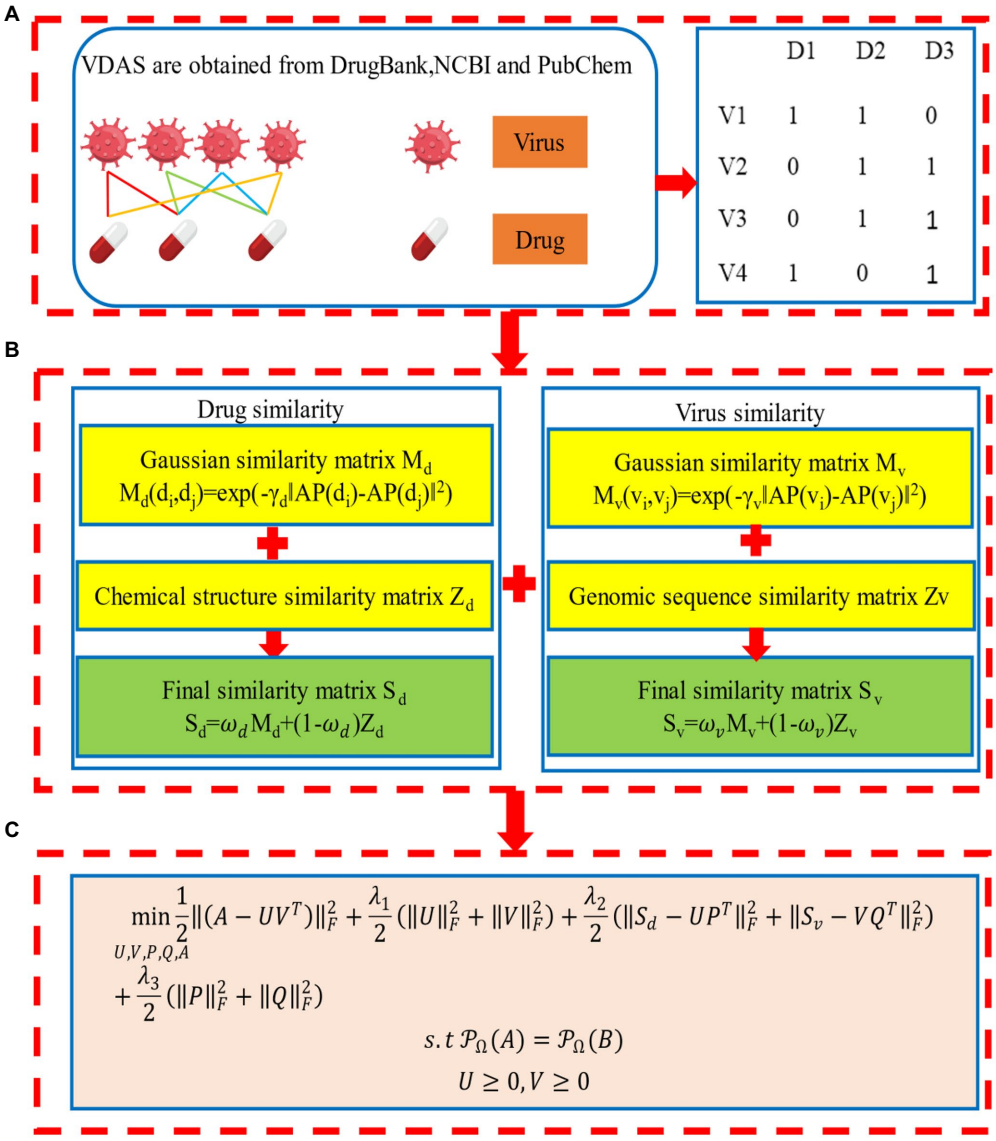

FIGURE 1
Workflow of Gaussian kernel similarity bilinear matrix factorization (VDA-GKSBMF). **(A)** Virus–drug association network and its association matrix. **(B)** Drug–drug similarity matrix and Virus–virus similarity matrix. **(C)** The model of VDA-GKSBMF.

TABLE 1 The statistics of datasets.

Datasets	No. of viruses	No. of drug	No. of VDAS	Sparsity
Dataset1	12	78	96	89.7%
Dataset2	69	128	770	91.3%
Dataset3	34	203	407	95.0%
Dataset4	34	210	437	93.9%
Dataset5	34	219	455	93.9%

Methods

Drug similarity matrix

Considering that drugs with common associated viruses may be similar, we denote the Gaussian association profile (AP) of drug d_i by $\mathrm{AP}(d_i)$, i.e., the i-th row of the VDA matrix B, which is a binary vector encoding the associations between this drug and viruses in the VDA matrix. Then, we calculate the similarity $M_d(d_i,d_j)$ between two drugs d_i and d_j based on association profiles of drugs by,

$$M_d(d_i,d_j)=\exp\left(-\gamma_d\|\mathrm{AP}(d_i)-\mathrm{AP}(d_j)\|^2\right)$$

where $\gamma_d=\gamma'_d/(\frac{1}{m}\sum_{k=1}^{m}\|AP(d_k)\|^2)$ is the normalized core band-width based on bandwidth parameter γ'_d, and m denotes the number of drugs.

Then, we obtain the chemical structure (CS)-based similarity between drugs calculated by RDKit (Landrum, 2014), which is

denoted as Z_d. Finally, we generate the drug–drug similarity matrix (DDS) by,

$$S_d = \omega_d M_d + (1-\omega_d) Z_d,$$

where $\omega_d \in [0,1]$ balances the contribution of the CS-based and AP-based drug similarity matrices. This forms a drug–drug network with edges weighted by the pairwise drug similarity scores.

Virus similarity matrix

Considering that viruses with common associated drugs may be similar, in the same way, we denote the Gaussian association profile (AP) of virus v_a by $\mathrm{AP}(v_a)$, i.e., the a -th column of the VDA matrix B, which is a binary vector encoding the associations between this virus and drugs in the VDA matrix. We calculate the AP-based similarity $M_v(v_a,v_b)$ between two viruses by,

$$M_v(v_a,v_b) = \exp\left(-\gamma_v \| \mathrm{AP}(v_a) - \mathrm{AP}(v_b) \|^2\right),$$

where $\gamma_v = \gamma'_v / (\frac{1}{n}\sum_{k=1}^{n} \| AP(v_k) \|^2)$, and n denotes the number of viruses.

Then, we obtain the sequence (SQ)-based similarity matrix calculated by MAFFT (Katoh and Toh, 2008), which is denoted as Z_v. Finally, the virus-virus similarity matrix (VVS) is calculated by,

$$S_v = \omega_v M_v + (1-\omega_v) Z_v,$$

where $\omega_v \in [0,1]$ balances the contribution of the SQ-based and AP-based virus similarity matrices. This forms a virus-virus network with edges weighted by the pairwise virus similarity scores.

Constructing heterogeneous network

To make use of information in the above DDS, VVS, and VDA matrices, we integrate them to construct a heterogeneous virus–drug network, by connecting the virus–virus network and drug–drug network through virus–drug associations. In the heterogeneous network, there are a set of m viruses $V = \{v_1, v_2, v_3, \ldots, v_m\}$ and a set of n drugs $D = \{d_1, d_2, d_3, \ldots, d_n\}$; the edge between drugs (d_i, d_j) is weighted by the score $S_d(d_i, d_j)$ in the DDS matrix, the edge between viruses (v_a, v_b) is weighted by the score $S_v(v_a, v_b)$ in the VVS matrix, and the edge between drug d_i and virus v_a denotes the existence of association between them.

The VDA matrix B is extremely sparse due to the rarity of known virus–drug associations, where 1/0 denotes known/unknown virus–drug associations, respectively. We would like to fill out the missing values in the matrix as scores to predict unknown VDAs. The integration of information of DDSs, VVSs, and known VDAs into the heterogeneous network will benefit the discovery of unknown VDAs due to the intrinsic correlation among drugs and viruses.

VDA-GKSBMF model to predict virus–drug associations

To predict potential virus-drug associations of COVID-19, we define the VDA prediction as a problem of completing virus-drug matrix in a heterogeneous virus-drug network, and explore potential VDAs of COVID-19 by Gaussian kernel similarity bilinear matrix factorization (Yang M. et al., 2020; called as VDA-GKSBMF).

Matrix factorization is an effective method, which intends to calculate an optimal approximation to the target matrix by decomposing it into two low-rank matrices. In a word, the mathematical model of matrix factorization is formulated as

$$\min_{U,V} B - UV^T{}_F^2, \quad (1)$$

where $B \in \mathbb{R}^{n\times m}$ is the given incomplete matrix with n drugs and m viruses, $\mathrm{U} \in \mathbb{R}^{n\times k}$ and $\mathrm{V} \in \mathbb{R}^{m\times k}$ are the indicator feature matrices of B and k is the subspace dimensionality [$\mathrm{k} \ll \min(\mathrm{n},\mathrm{m})$], $\|\cdot\|_F$ denotes the Frobenius norm. Many algorithms have been designed to provide numerical solutions for the above model or alternative forms. However, compared with other algorithms, the classic ADMM algorithm is superior to solving our proposed matrix factorization model.

The elements in the association matrix B are either 0 or 1. Thus, the predicted values in the un-known entries are expected to be in the interval of [0, 1], where a predicted value closer to 1 indicates that this is likely to be an indication and vice versa. Nevertheless, in the above matrix completion model, the entries in the completed matrix can be any real value in $(-\infty, +\infty)$.

Moreover, based on the assumption that similar drugs share similar molecular pathways to treat similar viruses, the underlying factors that determine drug-virus associations are highly correlated. Since B is extremely rare and low rank, usually less than 1% of known associations are present, while the rest of the elements are unknown. Therefore, the error term is only computed on items with known associations. At the same time, Tikhonov regularization terms are often used to avoid overfitting. To achieve this, the matrix factorization model can be expressed as,

$$\min_{U,V} \frac{1}{2}\mathcal{P}\Omega \| \left(B - UV^T\right) \|_F^2 + \frac{\lambda_1}{2}\left(\| U \|_F^2 + \| V \|_F^2\right), \quad (2)$$

where Ω is a set containing index pairs (i,j) of all known entries in B and $\mathcal{P}_\Omega$ is the projection operator onto Ω, λ_1 is regularization parameter. However, the above objective function does not involve a large amount of prior information about viruses and drugs, such as disease similarity and drug similarity. Since U and V are matrices containing potential eigenvectors of drugs and viruses, given a drug similarity matrix Z_d and a virus similarity matrix Z_v, UU^T and VV^T are expected to match S_d and S_v, respectively. Therefore, model (2) is described as follows:

$$\min_{U,V} \frac{1}{2}\| \mathcal{P}_{\Omega}\left(B - UV^T\right)\|_F^2 + \frac{\lambda_1}{2}\left(\| U \|_F^2 + \| V \|_F^2\right) + \frac{\lambda_2}{2}\left(\| Z_d - UU^T \|_F^2 + \| Z_v - VV^T \|_F^2\right) \quad (3)$$

Model (3) deals with a single drug and virus similarity measure. Here, in order to integrate the Gaussian kernel similarity measure, we propose the VDA-GKSBMF model, which is expressed as follows:

$$\min_{U,V,P,Q,A} \frac{1}{2}\| \left(A - UV^T\right)\|_F^2 + \frac{\lambda_1}{2}\left(\| U \|_F^2 + \| V \|_F^2\right) + \frac{\lambda_2}{2}\left(\| S_d - UP^T \|_F^2 + \| S_v - VQ^T \|_F^2\right) + \frac{\lambda_3}{2}\left(\| P \|_F^2 + \| Q \|_F^2\right) \quad (4)$$

$$s.t\ \mathcal{P}_{\Omega}(A) = \mathcal{P}_{\Omega}(B)$$

$$U \geq 0, V \geq 0,$$

where S_d and S_v are matrices concatenating Gaussian kernel similarity measure of drug and virus, and λ_1, λ_2, and λ_3 are balancing parameters. A is an auxiliary matrix for facilitating optimization. The approximation of similarity matrix S_d and S_v are constructed based on characteristic matrices U and V, where P and Q are potential characteristic matrices representing drug similarity and virus similarity, respectively. We solve model (4) by ADMM framework. Introducing two riving matrices X and Y, model (4) is transformed into

$$\min_{U,V,P,Q,X,Y,A} \frac{1}{2}\| \left(A - UV^T\right)\|_F^2 + \frac{\lambda_1}{2}\left(\| U \|_F^2 + \| V \|_F^2\right) + \frac{\lambda_2}{2}\left(\| S_d - UP^T \|_F^2 + \| S_v - VQ^T \|_F^2\right) + \frac{\lambda_3}{2}\left(\| P \|_F^2 + \| Q \|_F^2\right) \quad (5)$$

$$s.t\ \mathcal{P}_{\Omega}(A) = \mathcal{P}_{\Omega}(B)$$

$$U = X, V = Y$$

$$X \geq 0, Y \geq 0.$$

The augmented Lagrangian function becomes

$$\mathrm{L} = \| \left(A - UV^T\right)\|_F^2 + \frac{\lambda_1}{2}\left(\| U \|_F^2 + \| V \|_F^2\right) + \frac{\lambda_2}{2}\left(\| S_d - UP^T \|_F^2 + \| S_v - VQ^T \|_F^2\right) + \frac{\lambda_3}{2}\left(\| P \|_F^2 + \| Q \|_F^2\right) + Tr\left(W^T(U - X)\right) + Tr\left(R^T(U - X)\right) + \frac{\rho}{2}\left(\| U - X \|_F^2 + \| V - Y \|_F^2\right) \quad (6)$$

where W and R are the Lagrange multiplier and $\rho > 0$ is the penalty parameter. At the i-th iteration, it requires alternatively computing $U_{i+1}, V_{i+1}, P_{i+1}, Q_{i+1}, X_{i+1}, Y_{i+1}, A_{i+1}$.

Molecular docking method

Molecular docking method can be used to study the behavior of small molecules at the binding sites of target proteins. It has been widely used in drug design, since structures of more and more target proteins have been confirmed by experiments. AutoDock (Goodsell, 1996) is an open source molecular simulation software available to identify the conformation of a small molecule binding to a large molecule target. AutoDock has an affinity scoring function, which can sort candidate poses according to the sum of van der Waals and electrostatic energy. We used AutoDock to evaluate the molecular binding activity between predicted antiviral drugs and biomolecules.

Evaluation metrics

In this work, we evaluate the predictive performance of our method by 5-fold cross-validation. Popular evaluation metrics: AUC and AUPR are used to quantify the predictive performance of methods. Given a threshold of predictive scores, the candidate associations above this threshold are regarded as positives, and others are negatives. Then, true positive rate (TPR), false positive rate (FPR) and Precision can be calculated by,

$$\mathrm{TPR} = \mathrm{TP}/(\mathrm{TP}+\mathrm{FN}) \quad (7)$$

$$\mathrm{FPR} = \mathrm{FP}/(\mathrm{FP}+\mathrm{TN}) \quad (8)$$

$$\mathrm{Precision} = \mathrm{TP}/(\mathrm{TP}+\mathrm{FP}) \quad (9)$$

where TP, FP, TN, and FN represent true positive, false positive, true negative, and false negative, respectively. TPR is also called as Recall, which measures the ratio of correctly predicted positive samples to all positive samples. Precision measures the ratio of correctly predicted positive samples to all predicted positive samples.

With the increases of the threshold, TPR/Recall, FPR, and Precision will vary. TPR and FPR can form a TPR- FPR curve, called as the receiver-operating characteristic (ROC) curve. The area under the ROC curve is generally denoted as AUC. Precision and Recall (equivalent to TPR) can form a Precision–Recall (PR) curve. The area under the PR curve is generally denoted as AUPR. AUC and AUPR are scalar with the evaluation criterion: the larger AUC/AUPR is, the better the predictive performance is. AUC and AUPR can evaluate the overall performance of prediction algorithms.

Results

Parameter setting

In VDA-GKSBMF algorithm, there are tunable parameters $\gamma', \omega, \lambda_1, \lambda_2$ and λ_3. In order to prevent

multi-parameter overfitting, we set λ_1, λ_2 and λ_3 to the same value and remove two parameters. Because they are used to punish the related terms of U and V, P and Q in model (3) and model (4). VDA-GKSBMF has three parameters ($\gamma', \omega, \lambda_l$) needed to be determined. We first set γ' to 0.5, and then ω, λ_l are set in range of {0, 0.1, 0.2,…, 1}, {0.001, 0.01, 0.1, 1} by using the fivefold cross-validation on the training dataset. Table 2 displays the top 3 AUCS values as a function of $\gamma', \omega, \lambda_1, \lambda_2$ and λ_3 in five datasets.

Comparison with other methods

By 5-fold cross-validation experiment, we evaluate the performance of VDA-GKSBMF. We plot its ROC curve in Figure 2, and we find that it has a high AUC value in five datasets.

Further, we compare the VDA-GKSBMF method with other methods for drug repositioning: VDA-KATZ (Yang et al., 2019), IRNMF (Tang et al., 2020), VDA-GBNNR (Wang et al., 2021), and SCPMF (Meng et al., 2021). VDA-KATZ (Yang et al., 2019) used a KATZ algorithm to infer drug-virus association. The Indicator Regularized non-negative Matrix Factorization (IRNMF) method (Tang et al., 2020) introduced the indicator matrix and Karush-Kuhn-Tucker condition into the non-negative matrix factorization algorithm. VDA-GBNNR based on kernel similarity to predict anti-SARS-COV-2 drug. SCPMF used similarity constrained probabilistic matrix to infer drug-virus association. The experiment was carried out 50 times, with average performance as the final result. Table 3 shows sensitivities, specificities, accuracies, and AUCs of the five models on the five datasets. From Table 3, VDA-GBNNR obtains the best performance for other methods in dataset 1. However, VDA-GKSBMF achieves the best sensitivity, accuracy, specificity, and AUC on dataset 2, dataset 3, dataset 4, and dataset 5. Figure 2 displays the results of the methods in five datasets. The results show that the VDA-GKSBMF method outperforms the baseline methods in terms of the ROC curves and the corresponding AUC values, meaning that it can better discover antiviral drugs.

TABLE 2 The top three AUCs using different $\gamma', \omega, \lambda_1, \lambda_2$, and λ_3 values in 5-fold cross-validation.

Dataset	γ'	ω	λ_1	λ_2	λ_3	AUC
Dataset1	0.5	0.3	1	1	1	**0.8851**
	0.5	0.4	1	1	1	0.8825
	0.5	0.5	1	1	1	0.8663
Dataset2	0.5	0.1	0.1	0.1	0.1	**0.8594**
	0.5	0.2	0.1	0.1	0.1	0.8590
	0.5	0.3	0.1	0.1	0.1	0.8583
Dataset3	0.5	0.4	1	1	1	**0.8807**
	0.5	0.3	1	1	1	0.8793
	0.5	0.2	1	1	1	0.8756
Dataset4	0.5	0.2	0.1	0.1	0.1	**0.8824**
	0.5	0.3	0.1	0.1	0.1	0.8809
	0.5	0.4	0.1	0.1	0.1	0.8766
Dataset5	0.5	0.4	1	1	1	**0.8804**
	0.5	0.3	1	1	1	0.8789
	0.5	0.5	1	1	1	0.8787

Bold represented the best AUC values of different parameters in the same datasets.

Case study

After verifying the good performance of VDA-GKSBMF, to discover unknown antiviral drugs against SARS-CoV-2, we predict potential associations between SARS-CoV-2 and small molecule drugs based on known drug-virus association data, and we obtain the top-10 drugs with the highest score (see Table 4) in five datasets. Among the top-10 predicted drugs, there are 10 drugs that have been reported in the relevant literature, but the small molecule drugs were never confirmed to be anti-SARS-CoV-2 antiviral drugs. Ribavirin, Remdesivir, Oseltamivir, and Zidovudine were existed in at least four datasets.

Ribavirin is a road-spectrum antiviral drug that can inhibit the replication of respiratory syncytial virus (van Laarhoven and Marchiori, 2013). It can prevent respiratory syncytial virus infection in lung transplant recipients, and has been used to treat SARS-CoV and MERS-CoV. Similar to SARS-CoV and MERS-CoV, SARS-CoV-2 are a respiratory syndrome beta coronavirus that may cause severe respiratory diseases, and a few studies have reported that ribavirin may take an inhibitory effect on SARS-CoV-2 (Peng et al., 2020).

Remdesivir is a nucleoside analog with antiviral activity. Remdesivir has broad-spectrum activities against RNA viruses, such as SARS and MERS, and has been studied in a clinical trial for Ebola.

Oseltamivir is an antiviral neuraminidase inhibitor (Oseltamivir, n.d.) and has been used to prevent the infection of influenza A virus (for example, A-H1N1; Meijer et al., 2009, A-H5N1; De Jong et al., 2005, and influenza B virus). Oseltamivir can prevent the germination, replication, and infectivity of the virus in the host cell. More importantly, Oseltamivir combined with other drugs has been reported to inhibit the infection of SARS-CoV-2 (Huang et al., 2020).

Molecular docking

To further study the effectiveness of predicted drugs against SARS-CoV-2, the top 10 predicted small molecules are molecularly docked with SARS-CoV-2 spike protein/ACE2. From the DrugBank database, the chemical structures of these small molecule drugs have been obtained. The structure of spinous process protein of SARS-CoV-2 is calculated based on the homology model of Zhang lab (Wang et al., 2020). We used AutoDock, a bioinformatics tool, to conduct molecular docking between the predicted antiviral drug and SARS-CoV-2 spike

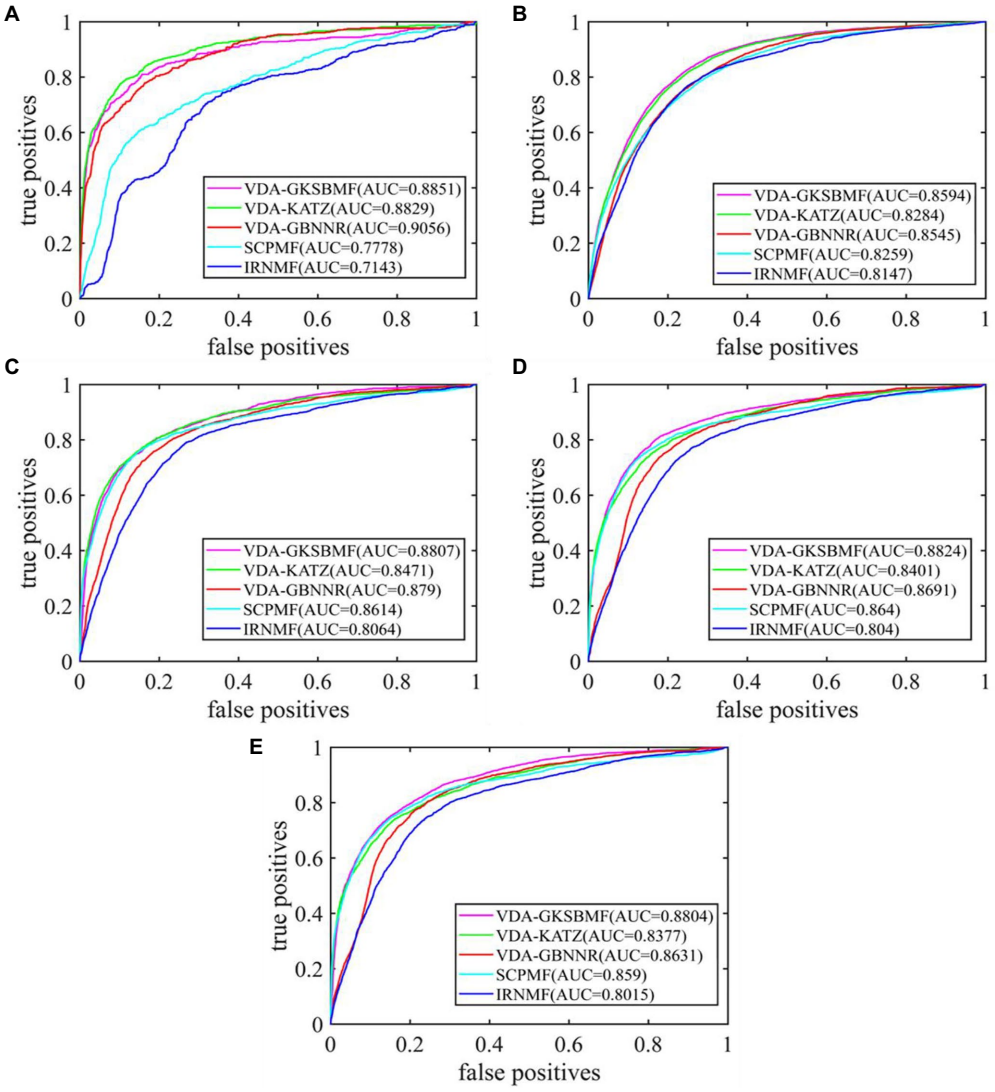

FIGURE 2
The performance of all methods in predicting virus–drug associations on five datasets: **(A)** Dataset1, **(B)** Dataset2, **(C)** Dataset3, **(D)** Dataset4, and **(E)** Dataset5.

protein/ACE2. The search algorithm scans the entire protein in AutoDock by genetic algorithm and grid box.

We calculate the predicted molecular binding energies of ribavirin, remdesivir, oseltamivir, and zidovudine small molecules with the spinous process protein and ACE2 of SARS-CoV-2 in Table 5. The results show that the binding activities of ribavirin with these two proteins are −5.29 and −6.39 kcal/mol, followed by remdesivir with −5.22 and −7.4 kcal/mol, and oseltamivir with −4.04 and − 4.73 kcal/mol. More importantly, ribavirin and remdesivir have been used to treat SARS, and their sequence homology with SARS-CoV-2 is about 79%.

Zidovudine has molecular binding energies of −6.54 and − 7.93 kcal/mol. Zidovudine is the drug which is an effective HIV replication inhibitor, which can improve immune function and partially reverse the neurological dysfunction caused by HIV. zidovudine, as an HIV nucleoside/nucleotide analogues reverse transcriptase inhibitor, has the potential to be a clue for SARS-COV-2 treatment.

Figures 3, 4 represent the docking results of four small molecules including ribavirin, remdesivir, oseltamivir, and zidovudine with two target proteins. The circles in each subgraph indicate the binding sites of the drug to the target protein. For example, the amino acids L387, L368, P565, and V209 are inferred to be the key residues for ribavirin binding to the SARS-CoV-2 spike protein/ACE2, while L849, T827, W1212, L144, and P504 are predicted as the key residues for remdesivir binding to these two target proteins.

TABLE 3 Performance indicators for different models.

Datasets	Methods	Accuracy	Sensitivity	Specificity	AUC
Dataset1	VDA-GKSBMF	0.5172	0.8757	0.5091	0.8851
	VDA-GBNNR	**0.5181**	**0.8957**	**0.5095**	**0.9056**
	VDA-KATZ	0.5171	0.8735	0.5090	0.8829
	SCPMF	0.5126	0.7708	0.5067	0.7778
	IRNMF	0.5098	0.7088	0.5052	0.7142
Dataset2	VDA-GKSBMF	**0.5136**	**0.8515**	**0.5072**	**0.8594**
	VDA-GBNNR	0.5134	0.8466	0.5071	0.8544
	VDA-KATZ	0.5125	0.8211	0.5066	0.8284
	SCPMF	0.5124	0.8187	0.5065	0.8259
	IRNMF	0.5120	0.8077	0.5063	0.8146
Dataset3	VDA-GKSBMF	**0.5097**	**0.8748**	**0.5052**	**0.8807**
	VDA-GBNNR	0.5097	0.8731	0.5051	0.8790
	VDA-KATZ	0.5089	0.8416	0.5047	0.8471
	SCPMF	0.5093	0.8557	0.5049	0.8613
	IRNMF	0.5079	0.8015	0.5042	0.8063
Dataset4	VDA-GKSBMF	**0.5102**	**0.8763**	**0.5054**	**0.8824**
	VDA-GBNNR	0.5098	0.8631	0.5052	0.8691
	VDA-KATZ	0.5091	0.8345	0.5048	0.8400
	SCPMF	0.5097	0.8581	0.5051	0.8639
	IRNMF	0.5081	0.7990	0.5044	0.8040
Dataset5	VDA-GKSBMF	**0.5101**	**0.8743**	**0.5054**	**0.8804**
	VDA-GBNNR	0.5096	0.8572	0.5051	0.8630
	VDA-KATZ	0.5090	0.8322	0.5048	0.8376
	SCPMF	0.5095	0.8532	0.5051	0.8590
	IRNMF	0.5081	0.7966	0.5043	0.8015

Bold represented the best value of different methods under the same evaluation condition.

TABLE 4 The predicted top-10 antiviral drugs against SARS-CoV-2 in five datasets.

Dataset1-drug	Dataset2-drug	Dataset3-drug	Dataset4-drug	Dataset5-drug
Remdesivir	Favipiravir	**Ribavirin**	Nitazoxanide	**Ribavirin**
Oseltamivir	**Remdesivir**	Nitazoxanide	**Ribavirin**	Chloroquine
Zanamivir	Cidofovir	Chloroquine	**Oseltamivir**	**Zidovudine**
ribavirin	**ribavirin**	Camostat	Camostat	Camostat
Laninamivir	Mycophenolic acid	Umifenovir	**Zidovudine**	Umifenovir
Peramivir	Navitoclax	**Remdesivir**	Favipiravir	Favipiravir
Presatovir	Itraconazole	**Zidovudine**	Hexachlorophene	Rifamycin
zidovudine	BCX4430 (Galidesivir)	Berberine	**Remdesivir**	**Oseltamivir**
Mycophenolic acid	Pleconaril	Amantadine	Sirolimus	Berberine
Mizoribine	Cyclosporine	**Oseltamivir**	Suramin	Niclosamide

Bold indicated that the drug existed in at least four datasets.

TABLE 5 The molecular binding energies between the predicted 4 antiviral drugs and two target proteins at least four datasets.

Drugs	Binding energies of target proteins	
	Spike protein	ACE2
Ribavirin	−5.29	−6.39
Remdesivir	−5.22	−7.40
Oseltamivir	−4.04	−4.73
Zidovudine	−6.54	−7.93

Discussion

Severe acute respiratory syndrome coronavirus 2 is quickly diffusing throughout the world, and it is urgent to find effective treatments against this virus. Drug repositioning, seeking to find new uses, offers a new strategy for the treatment of SARS-COV-2. However, to date, only a few databases have collated relevant drugs that may be used to treat SARS-COV-2. Thus, we developed a drug-virus as well

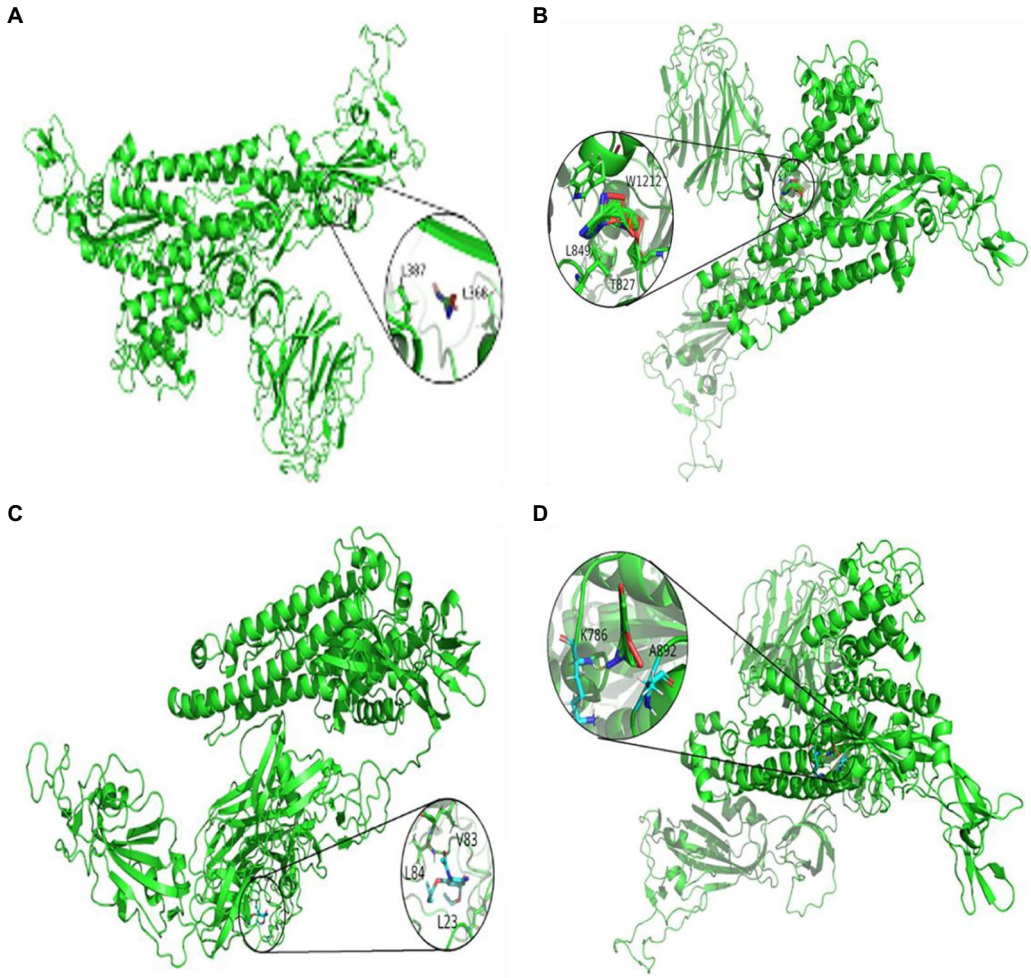

FIGURE 3
Molecular docking between the spike protein and four drugs: **(A)** ribavirin, **(B)** remdesivir, **(C)** oseltamivir, and **(D)** zidovudine.

as a method VDA-GKSBMF to prioritize drugs against SARS-COV-2.

Specifically, VDA-GKSBMF has a high AUC in cross-validation, which is better than other state-of-art methods in four datasets. We measured the molecular binding activity between predicted antiviral drugs and SARS-CoV-2 spike protein/human ACE2 (Zhao et al., 2020). Among them, the molecular binding energies between ACE2 and the four drugs were: Ribavirin (−6.39 kcal/mol), Remdesivir (−7.4 kcal/mol), Oseltamivir (−4.73 kcal/mol), zidovudine (−7.93 kcal/mol), and the four drugs have been in clinical trials or supported in recent publications. The results suggest that the VDA-GKSBMF algorithm can effectively infer unknown drugs of SARS-COV-2.

However, there a few limitations of this study. First, due to the limited size of the current virus-drug dataset and the complexity of intrinsic relationship in biomedical data, VDA-GKSBMF still has room for further improvement. On the one hand, we would like to expand the virus-drug dataset by including more virus-related and drug-related information, so as to further improve the predictive power of mining hidden virus-drug associations. On the other hand, it is also possible to enhance the ability of discovering potential drugs against SARS-COV-2 by more advanced and methods in related fields (Xu et al., 2020b; Xiang et al., 2021b, 2022a; Meng et al., 2022). Second, though we performed literature mining and molecular docking to validate our results, they are all in-silico methods. The prioritized drugs should be validated using wet-lab experiments. However, it is out of the scope of this study.

Conclusion

In this study, we collected five virus-drug datasets including VDAs matrix, virus genomic sequence similarity matrix, and drug chemical structure similarity matrix and explored drug repositioning of SARS-COV-2 by a novel method called VDA-GKSBMF.VDA-GKSBMF combined Gaussian similarity and extracted useful features to deduce potential virus-drug

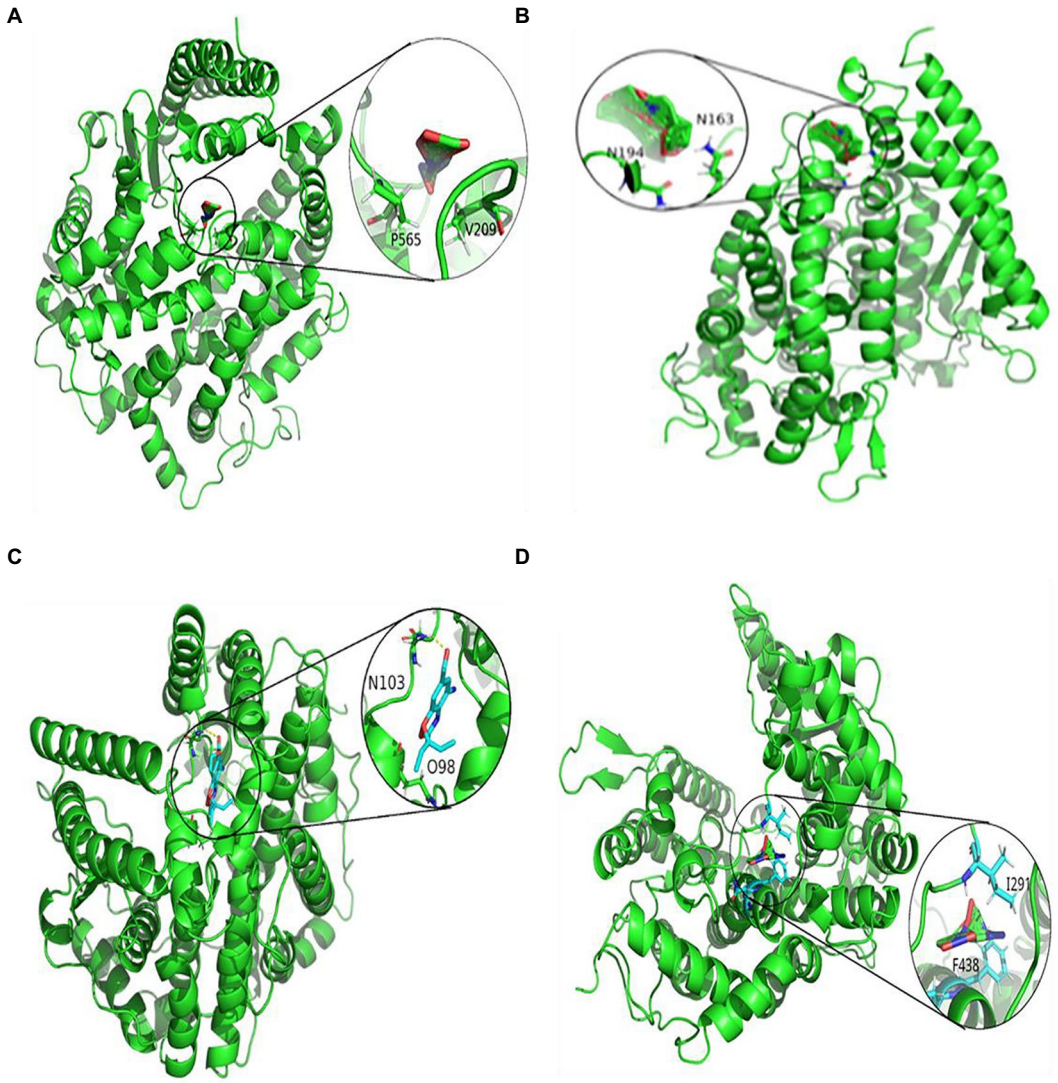

FIGURE 4
Molecular docking between ACE2 and four drugs: **(A)** ribavirin, **(B)** remdesivir, **(C)** oseltamivir, and **(D)** zidovudine.

associations. It combined Gaussian similarity and virus-drug association into the target function. The non-negative constraint was used in VDA-GKSBMF, ensuring that the predicted scores of association matrix were non-negative for the biological interpretability. Our results showed that VDA-GKSBMF is an effective approach for discovering new drugs of SARS-COV-2. In the future, we will combine different data resources to create larger dataset and design integrated algorithm, integrating multiple heterogeneous network and multiple similarities for predicting potential virus-drug associations.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found at: https://github.com/xiangju0208/VDA_GMSBMF.

Author contributions

BH and JH contributed to conception and design of the study. YW and JX organized the data and the prediction model. MT, RH, CL, and GT performed the statistical analysis. YW, JX, MB, JH, and BH wrote the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by the Training Program for Excellent Young Innovators of Changsha (Grant Nos. kq1802024, kq1905045, kq2009093, and kq2106075), Hunan key laboratory cultivation base of the research and development of novel pharmaceutical preparations (No. 2016TP1029), Hunan Provincial Innovation Platform and Talents Program (No. 2018RS3105), the Foundation of Hunan Educational Committee (Grant No. 19A060), and the

Provincial key R & D projects of Hunan Provincial Science and Technology Department (No. 2022SK2074). This research was funded by the Natural Science Foundation of Hunan province (No. 2018JJ2461), the Project to Introduce Intelligence from Oversea Experts to Changsha City (Grant No. 2089901), and General project of Education Department of Hunan Province (Grant No. 19C0190), and supported by the special fund of "Young and Middle-aged Key Teachers Training Program" of Changsha Medical College, the National Natural Science Foundation of China (32002235).

Conflict of interest

RH and GT are employed by Genesis (Beijing) Co. Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Cheng, L., Han, X., Zhu, Z., Qi, C., Wang, P., and Zhang, X. (2021a). Functional alterations caused by mutations reflect evolutionary trends of SARS-CoV-2. *Brief. Bioinform.* 22, 1442–1450. doi: 10.1093/bib/bbab042

Cheng, L., Zhu, Z., Wang, C., Wang, P., He, Y. O., and Zhang, X. (2021b). COVID-19 induces lower levels of IL-8, IL-10, and MCP-1 than other acute CRS-inducing diseases. *Proc. Natl. Acad. Sci. U. S. A.* 118:e2102960118. doi: 10.1073/pnas.2102960118

Cohain, A. T., Barrington, W. T., Jordan, D. M., Beckmann, N. D., Argmann, C. A., Houten, S. M., et al. (2021). An integrative multiomic network model links lipid metabolism to glucose regulation in coronary artery disease. *Nat. Commun.* 12:547. doi: 10.1038/s41467-020-20750-8

Coronaviridae Study Group of the International Committee on Taxonomy of V (2020). The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat. Microbiol.* 5, 536–544. doi: 10.1038/s41564-020-0695-z

De Jong, M. D., Tran, T. T., Truong, H. K., Vo, M. H., Smith, G. J., Nguyen, V. C., et al. (2005). Oseltamivir resistance during treatment of influenza a (H5N1) infection. *N. Engl. J. Med.* 353, 2667–2672. doi: 10.1056/NEJMoa054512

Eurosurveillance editorial team (2020). Note from the editors: World Health Organization declares novel coronavirus (2019-nCoV) sixth public health emergency of international concern. *Eur. Secur.* 25:200131e. doi: 10.2807/1560-7917.ES.2020.25.5.200131e

Goodsell, D. S. (1996). Automated docking of flexible ligands: Applications of autodock molecular recognition.

Gralinski, L. E. (2020). Menachery VD: return of the coronavirus: 2019-nCoV. *Viruses* 12:135. doi: 10.3390/v12020135

He, B., Wang, K., Xiang, J., Bing, P., Tang, M., Tian, G., et al. (2022). DGHNE: network enhancement-based method in identifying disease-causing genes through a heterogeneous biomedical network. *Brief. Bioinform.* 23:bbac405. doi: 10.1093/bib/bbac405

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., et al. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 395, 497–506. doi: 10.1016/S0140-6736(20)30183-5

Katoh, K., and Toh, H. (2008). Recent developments in the MAFFT multiple sequence alignment program. *Brief. Bioinform.* 9, 286–298. doi: 10.1093/bib/bbn013

Kim, S., Thiessen, P. A., Bolton, E. E., Chen, J., Fu, G., Gindulyte, A., et al. (2016). PubChem substance and compound databases. *Nucleic Acids Res.* 44, D1202–D1213. doi: 10.1093/nar/gkv951

Landrum, G. (2014). RDKit:open-source cheminformatics. Release 2014.03.1.

Li, J., Wang, X., Li, N., Jiang, Y., Huang, H., Wang, T., et al. (2020). Feasibility of mesenchymal stem cell therapy for COVID-19: a mini review. *Curr. Gene Ther.* 20, 285–288. doi: 10.2174/1566523220999200820172829

Liu, H., Qiu, C., Wang, B., Bing, P., Tian, G., Zhang, X., et al. (2021). Evaluating DNA methylation, gene expression, somatic mutation, and their combinations in inferring tumor tissue-of-origin. *Front. Cell Dev. Biol.* 9:619330. doi: 10.3389/fcell.2021.772380

Liu, C., Wei, D., Xiang, J., Ren, F., Huang, L., Lang, J., et al. (2020). An improved anticancer drug-response prediction based on an ensemble method integrating matrix completion and ridge regression. *Mol. Ther. Nucleic Acids* 21, 676–686. doi: 10.1016/j.omtn.2020.07.003

Liu, X., Yang, J., Zhang, Y., Fang, Y., Wang, F., Wang, J., et al. (2016). A systematic study on drug-response associated genes using baseline gene expressions of the cancer cell line encyclopedia. *Sci. Rep.* 6:22811. doi: 10.1038/srep22811

Lu, K., Wang, F., Ma, B., Cao, W., Guo, Q., Wang, H., et al. (2021). Teratogenic toxicity evaluation of bladder cancer-specific oncolytic adenovirus on mice. *Curr. Gene Ther.* 21, 160–166. doi: 10.2174/1566523220999201217161258

Meijer, A., Lackenby, A., Hungnes, O., Lina, B., Van-Der-Werf, S., Schweiger, B., et al. (2009). On behalf of the European influenza surveillance scheme: oseltamivir-resistant influenza virus a (H1N1), Europe, 2007-08 seasona. *Emerg. Infect. Dis.* 15, 552–560. doi: 10.3201/eid1504.181280

Meng, Y., Jin, M., Tang, X., and Xu, J. (2021). Drug repositioning based on similarity constrained probabilistic matrix factorization: COVID-19 as a case study. *Appl. Soft Comput.* 103:107135. doi: 10.1016/j.asoc.2021.107135

Meng, Y., Lu, C., Jin, M., Xu, J., Zeng, X., and Yang, J. (2022). A weighted bilinear neural collaborative filtering approach for drug repositioning. *Brief. Bioinform.* 23:bbab581. doi: 10.1093/bib/bbab581

Novac, N. (2013). Challenges and opportunities of drug repositioning. *Trends Pharmacol. Sci.* 34, 267–272. doi: 10.1016/j.tips.2013.03.004

Oseltamivir (n. d.). Oseltamivir: Description Available at: https://www.drugbank.ca/drugs/DB00198

Parsza, C. N., Gomez, D. L. M., Simonin, J. A., Nicolas Belaich, M., and Ghiringhelli, P. D. (2021). Evaluation of the Nucleopolyhedrovirus of Anticarsia gemmatalis as a vector for gene therapy in mammals. *Curr. Gene Ther.* 21, 177–189. doi: 10.2174/1566523220999201217155945

Peng, L., Tian, X., Shen, L., Kuang, M., Li, T., Tian, G., et al. (2020). Identifying effective antiviral drugs against SARS-CoV-2 by drug repositioning through virus-drug association prediction. *Front. Genet.* 11:577387. doi: 10.3389/fgene.2020.577387

Shen, L., Liu, F., Huang, L., Liu, G., Zhou, L., and Peng, L. (2022). VDA-RWLRLS: an anti-SARS-CoV-2 drug prioritizing framework combining an unbalanced bi-random walk and Laplacian regularized least squares. *Comput. Biol. Med.* 140:105119. doi: 10.1016/j.compbiomed.2021.105119

Tang, X., Cai, L., Meng, Y., Xu, J., Lu, C., and Yang, J. (2020). Indicator regularized non-negative matrix factorization method-based drug repurposing for COVID-19. *Front. Immunol.* 11:603615. doi: 10.3389/fimmu.2020.603615

van Laarhoven, T., and Marchiori, E. (2013). Predicting drug-target interactions for new drug compounds using a weighted nearest neighbor profile. *PLoS One* 8:e66952. doi: 10.1371/journal.pone.0066952

Wang, J., Wang, C., Shen, L., Zhou, L., and Peng, L. (2021). Screening potential drugs for COVID-19 based on bound nuclear norm regularization. *Front. Genet.* 12:817672. doi: 10.3389/fgene.2021.817672

Wang, F., Yang, J., Lin, H., Li, Q., Ye, Z., Lu, Q., et al. (2020). Improved human age prediction by using gene expression profiles from multiple tissues. *Front. Genet.* 11:1025. doi: 10.3389/fgene.2020.01025

Wheeler, D. L., Church, D. M., Edgar, R., Federhen, S., Helmberg, W., Madden, T. L., et al. (2004). Database resources of the National Center for biotechnology information: update. *Nucleic Acids Res.* 32, 35D–340D. doi: 10.1093/nar/gkh073

Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., et al. (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* 46, D1074–D1082. doi: 10.1093/nar/gkx1037

Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G., et al. (2020). A new coronavirus associated with human respiratory disease in China. *Nature* 579, 265–269. doi: 10.1038/s41586-020-2008-3

Xiang, J, Meng, X, Zhao, Y, Wu, F-X, and Li, M. (2022a). HyMM: hybrid method for disease-gene prediction by integrating multiscale module structure. *Brief. Bioinform.* doi: 10.1093/bib/bbac072 [Epub ahead of print].

Xiang, J., Zhang, N.-R., Zhang, J.-S., Lv, X.-Y., and Li, M. (2021a). PrGeFNE: predicting disease-related genes by fast network embedding. *Methods* 192, 3–12. doi: 10.1016/j.ymeth.2020.06.015

Xiang, J, Zhang, J, Zhao, Y, Wu, F-X, and Li, M. (2022b). Biomedical data, computational methods and tools for evaluating disease–disease associations. *Brief. Bioinform.* doi: 10.1093/bib/bbac006 [Epub ahead of print].

Xiang, J., Zhang, J., Zheng, R., Li, X., and Li, M. (2021b). NIDM: network impulsive dynamics on multiplex biological network for disease-gene prediction. *Brief. Bioinform.* 22:bbab080. doi: 10.1093/bib/bbab080

Xu, J., Cai, L., Liao, B., Zhu, W., and Yang, J. (2020a). CMF-impute: an accurate imputation tool for single-cell RNA-seq data. *Bioinformatics* 36, 3139–3147. doi: 10.1093/bioinformatics/btaa109

Xu, J., Zhu, W., Cai, L., Liao, B., Meng, Y., Xiang, J., et al. (2020b). LRMCMDA: predicting miRNA-disease association by integrating low-rank matrix completion with miRNA and disease similarity information. *IEEE Access* 8, 80728–80738. doi: 10.1109/ACCESS.2020.2990533

Yang, J., Ju, J., Guo, L., Ji, B., Shi, S., Yang, Z., et al. (2022). Prediction of HER2-positive breast cancer recurrence and metastasis risk from histopathological images and clinical information via multimodal deep learning. *Comput. Struct. Biotechnol. J.* 20, 333–342. doi: 10.1016/j.csbj.2021.12.028

Yang, M., Luo, H., Li, Y., and Wang, J. (2019). Drug repositioning based on bounded nuclear norm regularization. *Bioinformatics* 35, i455–i463. doi: 10.1093/bioinformatics/btz331

Yang, J., Peng, S., Zhang, B., Houten, S., Schadt, E., Zhu, J., et al. (2020). Human geroprotector discovery by targeting the converging subnetworks of aging and age-related diseases. *Geroscience* 42, 353–372. doi: 10.1007/s11357-019-00106-x

Yang, M., Wu, G., Zhao, Q., Li, Y., and Wang, J. (2020). Computational drug repositioning based on multi-similarities bilinear matrix factorization. *Brief. Bioinform.* 22:bbaa267. doi: 10.1093/bib/bbaa267

Yao, Y., Li, X., Liao, B., Huang, L., He, P., Wang, F., et al. (2017). Predicting influenza antigenicity from Hemagglutintin sequence data based on a joint random forest method. *Sci. Rep.* 7:1545. doi: 10.1038/s41598-017-01699-z

Zhang, Y., Huang, H., Zhang, D., Qiu, J., Yang, J., Wang, K., et al. (2017). A review on recent computational methods for predicting noncoding RNAs. *Biomed. Res. Int.* 2017:9139504. doi: 10.1155/2017/9139504

Zhang, Y., Xiang, J., Tang, L., Li, J., Lu, Q., Tian, G., et al. (2021). Identifying breast cancer-related genes based on a novel computational framework involving KEGG pathways and PPI network modularity. *Front. Genet.* 12:596794. doi: 10.3389/fgene.2021.809608

Zhang, W., Zhang, H., Yang, H., Li, M., Xie, Z., and Li, W. (2019). Computational resources associating diseases with genotypes, phenotypes and exposures. *Brief. Bioinform.* 20, 2098–2115. doi: 10.1093/bib/bby071

Zhao, Y, Zhao, Z, Wang, Y, Zhou, Y, Ma, Y, and Zuo, W. (2020). Single-cell RNA expression profiling of ACE2, the receptor of SARS-COV-2. *Am. J. Respir. Crit. Care. Med.* 202, 756–759. doi: 10.1164/rccm.202001-0179LE

Zhou, L., Wang, J., Liu, G., Lu, Q., Dong, R., Tian, G., et al. (2020). Probing antiviral drugs against SARS-CoV-2 through virus-drug association prediction based on the KATZ method. *Genomics* 112, 4427–4434. doi: 10.1016/j.ygeno.2020.07.044

Zhou, P., Yang, X. L., Wang, X. G., Hu, B., Zhang, L., Zhang, W., et al. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579, 270–273. doi: 10.1038/s41586-020-2012-7

Zhu, Z., Zhang, S., Wang, P., Chen, X., Bi, J., Cheng, L., et al. (2021). A comprehensive review of the analysis and integration of omics data for SARS-CoV-2 and COVID-19. *Brief. Bioinform.* 23:bbab446. doi: 10.1093/bib/bbab302

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., et al. (2020). A novel coronavirus from patients with pneumonia in China, 2019. *N. Engl. J. Med.* 382, 727–733. doi: 10.1056/NEJMoa2001017