Computers in Biology and Medicine 159 (2023) 106969

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

DTSEA: A network-based drug target set enrichment analysis method for drug repurposing against COVID-19

Yinchun Su [a,1], Jiashuo Wu [b,1], Xiangmei Li [b,1], Ji Li [b,1], Xilong Zhao [b], Bingyue Pan [b], Junling Huang [b], Qingfei Kong [a,**], Junwei Han [b,*]

[a] Department of Neurobiology, Harbin Medical University, Harbin, 150081, PR China
[b] College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, 150081, PR China

ARTICLE INFO

Keywords:
COVID-19
Drug repurposing
Network medicine
Network propagation approach
Enrichment analysis

ABSTRACT

The Coronavirus Disease 2019 (COVID-19) pandemic is still wreaking havoc worldwide. Therefore, the urgent need for efficient treatments pushes researchers and clinicians into screening effective drugs. Drug repurposing may be a promising and time-saving strategy to identify potential drugs against this disease. Here, we developed a novel computational approach, named Drug Target Set Enrichment Analysis (DTSEA), to identify potent drugs against COVID-19. DTSEA first mapped the disease-related genes into a gene functional interaction network, and then it used a network propagation algorithm to rank all genes in the network by calculating the network proximity of genes to disease-related genes. Finally, an enrichment analysis was performed on drug target sets to prioritize disease-candidate drugs. It was shown that the top three drugs predicted by DTSEA, including Ataluren, Carfilzomib, and Aripiprazole, were significantly enriched in the immune response pathways indicating the potential for use as promising COVID-19 inhibitors. In addition to these drugs, DTSEA also identified several drugs (such as Remdesivir and Olumiant), which have obtained emergency use authorization (EUA) for COVID-19. These results indicated that DTSEA could effectively identify the candidate drugs for COVID-19, which will help to accelerate the development of drugs for COVID-19. We then performed several validations to ensure the reliability and validity of DTSEA, including topological analysis, robustness analysis, and prediction consistency. Collectively, DTSEA successfully predicted candidate drugs against COVID-19 with high accuracy and reliability, thus making it a formidable tool to identify potential drugs for a specific disease and facilitate further investigation.

1. Introduction

On December 1, 2019, coronavirus disease 2019 (COVID-19) rapidly spread and became a global pandemic [1]. It is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [2], a novel human coronavirus strain associated with a long evolutionary history in a bat host and a suspected spillover event into humans [3]. Despite the World Health Organization declaring the pandemic a public health emergency in its early stages, the high fatality rate [4] and high transmissibility [5] of COVID-19 make it challenging to control its spread. The public has placed immense pressure on researchers and clinicians to develop effective prevention and treatment methods. However, it typically takes a decade to develop and mass-produce a novel drug for the current pandemic with the assurance that it is safe and effective [6]. This delay is unacceptable and exacerbates the pandemic.

Drug repurposing offers an alternative to the traditional drug development process, we can repurpose drugs for treating SARS-CoV-2 infection with existing drugs. The major advantage of drug repurposing is its efficiency, which is a key drawback of the traditional drug development process. For example, Remdesivir was initially developed to treat hepatitis C [7] and Ebola [8] infections, but it was later approved to treat COVID-19 in various countries [9]. It only takes one year for Remdesivir to be approved for marketing, and its safety is not a major concern because it has already passed most of its validation stages. Thus, the repurposed drugs have known safety and pharmacological profiles. Experimental and computational methods are used to screen potential

* Corresponding author.
** Corresponding author.
E-mail addresses: kqfangel@hrbmu.edu.cn (Q. Kong), hanjunwei@ems.hrbmu.edu.cn (J. Han).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.compbiomed.2023.106969
Received 13 December 2022; Received in revised form 27 March 2023; Accepted 19 April 2023
Available online 21 April 2023

novel indications for existing drugs. The latter approach relies on large-scale biological data to yield a more robust and reliable result than the former [10]. To combat COVID-19 using the latter approach, determining the most suitable and efficient drug repurposing algorithm has become a significant challenge.

Recently, several computational approaches have been used to assess the similarity between drugs and COVID-19 [11]. These approaches typically use chemical or protein structures, network-based algorithms, machine learning algorithms, and data mining techniques to identify novel potential treatment drugs. For example, machine-learning approaches are now underway to analyze massive amounts of public data to accurately predict the clinical outcomes of patients with COVID-19 [12]. Even though the results of machine-learning approaches are encouraging, it is worth noticing that interpreting the fitted models can be much more challenging than the traditional methods [13].

Among various kinds of computational approaches, the structure-based approaches are the ones of the most widely adopted methods for screening novel drugs. This approach analyzes the similarities between the binding site of SARS-CoV-2 (such as ACE2) and drug structures [14,15], leveraging the abundance of 3D structures of drugs and receptor targets [16,17]. To address the deficiency in computing similarity scoring matches, insertions, and deletions in sequence, the fuzzy logic-based computational method was proposed for achieving better predictive results in real-world applications [18]. However, targeting a single protein or gene oversimplifies disease progression and immune response, and also ignores the complex interactions between genes or proteins in biological systems.

To offset the problems mentioned above, the network-based approaches offer a holistic perspective by integrating multiple sources of information from a holistic perspective. In biological networks, nodes can represent various types of information (such as genes, pathways, or drugs) that are not independent entities [19–22]. Instead, the nodes are connected by specific link types (such as interactions or modulations) depending on the type of network. We posit that identifying genes affected by COVID-19 in the network can be more accurate and systematic than through differential analysis of individual genes alone.

In the study, we present a novel network-based drug repurposing method, named drug target set enrichment analysis (DTSEA), to identify candidate drugs against COVID-19 by integrating the network propagation algorithm and enrichment analysis strategy. DTSEA first uses the random walk with restart (RWR) algorithm to calculate the network proximity that is distant from the disease-related genes in the large-scale human gene functional interaction network. This allows us to estimate the influence of the disease and accurately and systematically identify genes affected by COVID-19. The method then predicts drug-disease associations by calculating the enrichment scores of the drug target sets in the ranked gene list based on the network proximity to determine whether a drug is potent in treating COVID-19. The algorithm sheds light on a novel metric for measuring drug-disease distance and provides a unique insight into drug repurposing for combating human disease. We have wrapped the core function into an R package named DTSEA (Drug Target Set Enrichment Analysis), which is freely available on CRAN under the GPL-v2 license (https://CRAN.R-project.org/package=DTSEA).

2. Materials and methods

2.1. Data preparation

2.1.1. Identification of the disease-related genes

We obtained the COVID-19 gene expression dataset from the Gene Expression Omnibus (GEO) database (ID: GSE183071) [23]. Because of local immune response could be key to determine the course of the systemic response and thus COVID-19 severity, the GSE183071 dataset focused on immune-related genes from patients affected by different COVID-19 severities. This collection comprises 156 samples from three different tissue types (blood, nasal, and saliva) among 77 distinct donors. Given that gene expression varies between tissues and is tissue-specific, we only extracted the nasal tissue samples from the dataset. Ultimately, we obtained 37 disease samples and 13 controls to identify disease-related genes.

We conducted Welch's *t*-test with subsequent Benjamini-Hochberg correction (or the false discovery rate, FDR) to identify differentially expressed genes (DEGs) as COVID-19-related genes by comparing gene expression levels between disease and control samples. To avoid false positives, we considered only genes with $FDR < 0.01$ as DEGs. Additionally, to obtain a comprehensive understanding of the disease landscape, we included 11 genes (TLR4, NLRP3, MBL2, IL6, IL1RN, IL1B, CX3CR1, CCR5, AGT, ACE, and F2) associated with both COVID-19 and its three comorbidities (hypertension, diabetes mellitus, and coronary artery disease) proposed by Feng et al. [24].

We then acquired three distinct gene expression datasets to validate our results, which were collected from the GEO database (ID: GSE156544, GSE164805, and GSE177477). The detailed information of the three datasets is provided in Supplementary Table S1. We also collected 190 COVID-19-related genes of human from NCBI in June 2022 (https://www.ncbi.nlm.nih.gov/gene/?term=coronavirus+related+%5Bproperties%5D).

2.1.2. Construction of a gene functional interaction network

We collected human pathways from seven databases, including KEGG, Reactome, Biocarta, NCI, SPIKE, HumanCyc, and Panther. We then constructed an undirected human gene functional interaction network by integrating the gene relationships in the pathways, compromising 221,353 functional interactions (edges) and 12,836 unique genes (nodes). This network was deposited in our DTSEAData package on GitHub (https://github.com/hanjunwei-lab/DTSEAData).

2.1.3. Collecting the drug-related information

The drug-related information used in this study includes: (1) clinical trial data of drugs, (2) drug medical indications, and (3) drug target information. We collected clinical trial data of drugs and drug indications from the ChEMBL database (Version CHEMBL30, March 2022) [25], which provides 14 kilos of verified drugs, drug clinical trial data, and medical indications. Next, we retrieved drug-target interactions from both the ChEMBL and the DrugBank (Version 5.1.9, Jan 2022) [26] database, which resulted in 5,804 drugs and 17,155 drug-target interactions. Moreover, we manually coded COVID-19-related symptoms (both Omicron variant and long COVID symptoms, as provided in Supplementary Table 2) from the ChEMBL database to validate the prediction results qualitatively.

2.2. Study design of the DTSEA framework

Our hypothesis posits that potential drug candidates for a specific disease (e.g., COVID-19) should be located near or in close proximity to the disease in the network. To determine whether a drug is effective against the chosen disease, we use DTSEA algorithm to assess the network proximity between drug targets and disease-related genes.

The proposed DTSEA method involves four main steps for identifying candidate drugs for a given disease: (1) Map the disease-related genes into the network of gene functional interaction network. (2) Compute the distance between disease-related genes and other gene nodes by the random walk with restart (RWR) algorithm, then rank the nodes decreasingly based on the RWR result. (3) Map the drug target set of each drug to the ranked list to calculate the drug target set enrichment score (ES) through the GSEA approach, and then the resulting drug-wise ES scores indicate the mean distance towards the disease per drug. (4) Perform permutation analysis and sort the result list by the normalized ES (NES) and FDR, and then the drug candidates are filtered by the manually-coded criteria. The schematic overview of the DTSEA method is shown in Fig. 1.

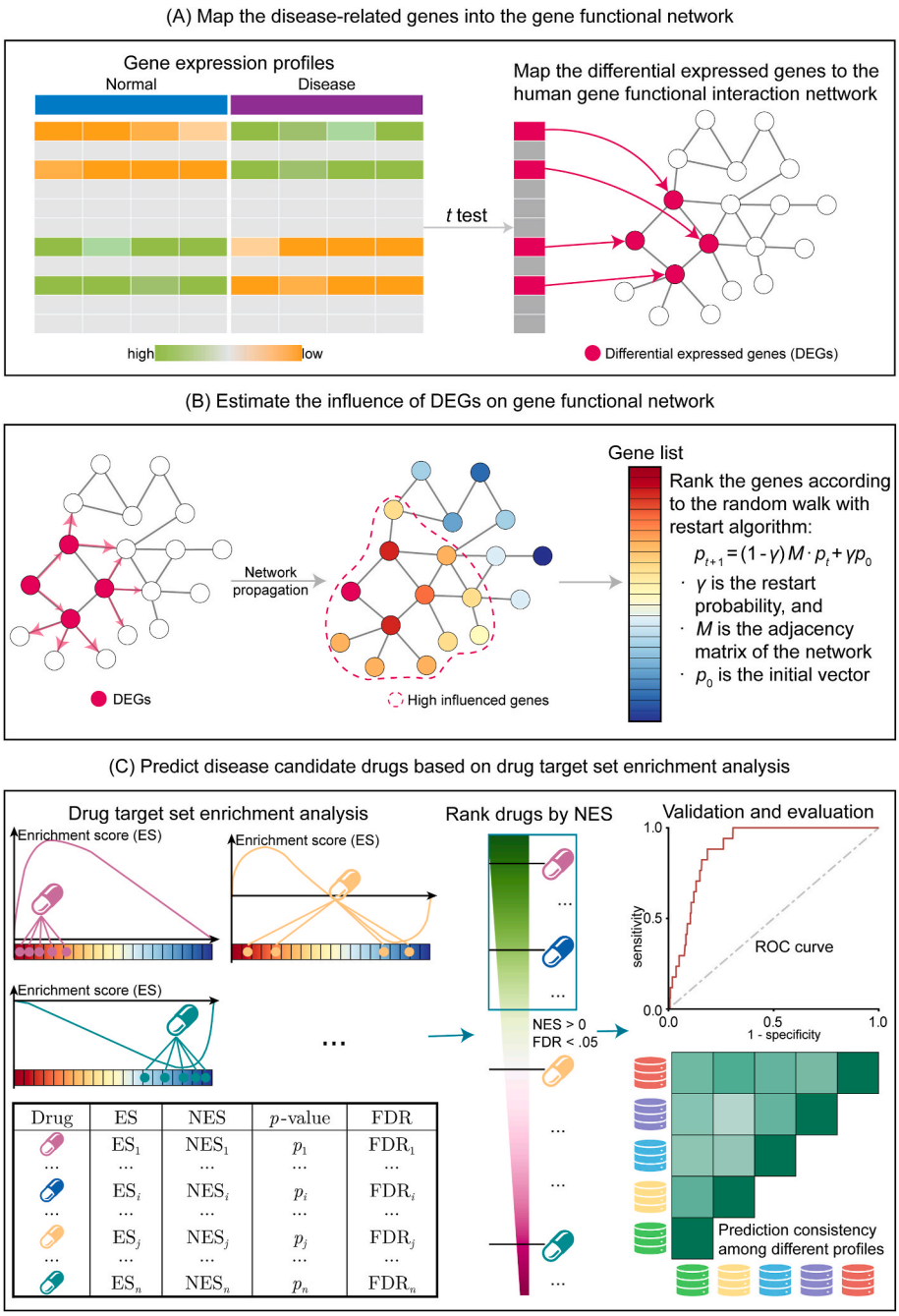

Fig. 1. A simple schematic of the DTSEA method. **(A)** Map the disease-related genes into the gene functional interaction network. The disease-related genes can be obtained from numerous publicly available databases, the Gene Expression Omnibus (GEO) database for expression profiles or the gene list from the disease-related gene database. After performing the *t*-test, the differentially expressed genes (DEGs) can be identified and treated as disease-related genes. **(B)** Estimate the influence of disease-related genes in the network. The RWR approach calculated the proximity between disease genes and other genes, and then ranked the nodes decreasingly based on the RWR result. **(C)** Predict disease candidate drugs based on drug target set enrichment analysis. The drug target sets were respectively mapped to the ranked gene list to calculate the drug target set ESs, and ranked the drugs by NES.

2.2.1. Assessing the proximity between disease genes and other genes in the network

RWR is the modified version of the standard random walk algorithm, which simulates an iterative random walker that starts from a set of source nodes and travels to its immediate neighbors, or returns to the source nodes at each time step in a graph. This algorithm could be used for quantifying the network proximity between any nodes in a network and a given set of nodes. Here, we adopted the RWR algorithm to assess the proximity between disease genes and other genes in the gene functional interaction network (Fig. 1B). We shall define the RWR process with n nodes (COVID-19-related genes) in the network as:

$$p^{t+1} = (1-\gamma)Mp^t + \gamma p^0 \quad (1)$$

where M is the column-normalized adjacency matrix with the network, $p^t = (p_1^t, p_2^t, \ldots, p_n^t)'$ is the visiting probability of each node at time step t, in which the i th element p_i^t represents the probability of node i at time step t. $p^0 = (p_1^0, p_2^0, \ldots, p_n^0)'$ represents the initial probability vector of nodes, where the nodes in the restart set corresponding to COVID-19 related genes are assigned as 1 and remaining nodes as 0, and which was then normalized to a unit vector. In this study, we set the restart probability $\gamma = 0.7$, because it had little effect on the RWR results between 0.1 and 0.9 [27]. After a certain number of iterations, the probability vector p^t will eventually arrive at a stable state when the difference between p^{t+1} and p^t decreases to less than 10^{-10}. The vector p^t was then normalized according to the median of its elements. A gene i in the network with a larger value p_i^t indicates it is more proximal the COVID-19-related gene nodes. A ranked gene list L was then constructed by descending p^t.

2.2.2. Calculating the enrichment score of drug target set

According to the DTSEA hypothesis, the effectiveness of a drug is maximized when its targets are within close proximity of the disease-related nodes in the network. To test this hypothesis, we respectively mapped the target genes of each drug to the ranked gene list L to determine whether target genes of a drug tend to occur toward the top of the list. In this case, the drug targets are close to the disease-related genes in the network, which may have a potential treatment effect on the disease. We used the weighted Kolmogorov-Smirnov (KS) statistics to calculate an *ES* of a drug target gene set that reflects the degree to which the drug target set is overrepresented at the top of the entire ranked list L. If the target genes of a drug cluster at the top of the list, the *ES* will be relative large ($ES > 0$). This statistic has been used in GSEA [28] previously to identify diseases related gene sets. In the paper, we used it as a statistic test of drug target gene set to prioritize disease candidate drugs.

To assess the statistical significance (normal *p-value*) of an observed *ES*, we performed permutation test. Specifically, for a given drug i, we randomly sampled gene sets of the same size as the drug target gene sets from the ranked gene list L, and re-computed *ES*. This permutation process was repeated n times (we set $n = 1000$ in the study), and an empirical null distribution ES_{null} was obtain. The normal p-value was then estimated by comparing the observed *ES* with ES_{null}, and which was then adjusted by using false discovery rate (FDR) method [29]. Simultaneously, the normalized ES (NES) for each drug was also calculated as performed in the GSEA method. We used the fgsea package [30] to implement the above process.

2.3. Measuring the performance of DTSEA

We evaluated the performance of DTSEA using several quantitative and qualitative approaches. We first screened the predicted top drugs based on their indications. Second, we computed the network distances between the drug targets of each candidate drug and disease genes, and compared them with the DTSEA predictions to validate the primary hypothesis of DTSEA. Third, we assessed the robustness of DTSEA by randomly removing a certain proportion of the nodes and edges. Furthermore, we applied the DTSEA to multiple COVID-19 datasets and examined its predictive consistency. Finally, we extended the DTSEA algorithm to multiple breast cancer datasets to assess its effectiveness.

2.3.1. The network separation metric

We employed the network separation metric s_{AB} to validate the extent of overlap between a potential drug target set (called module A) and disease-related gene sets (called module B). This metric compares the mean shortest distances between two modules and is defined as:

$$s_{AB} = d_{AB} - \frac{d_A + d_B}{2} \tag{2}$$

where d_{AB} is the mean shortest distance between module A and module B (between-module distance), and either d_A or d_B is the mean shortest distance within each module A and B (within-module distance). The shortest distance is the count or summed weight of all nodes in the most efficient path between two nodes. We can calculate the between-module distance d_{AB} by averaging the shortest distances from each node in module A to module B. The within-module distance d_A or d_B can be derived similarly but differ slightly. In calculating the distance d_A, we sequentially remove the nodes i from module A and compute the mean distance between i and module A excluding node i. Accordingly, the separation metric $s_{AB} < 0$ indicates network overlap, whereas $s_{AB} > 0$ indicates non-overlap [31].

2.3.2. The discrimination of the method

The receiver operating characteristic (ROC) curve was used to evaluate the accuracy of DTSEA. A ROC curve illustrates the trade-offs between true positive (benefits) and false positive (costs) [32], and the area under the curve (AUC) quantifies the trade-off. It is defined as:

$$AUC = \int_{x=0}^{1} TPR\left(FPR^{-1}(x)\right)dx \tag{3}$$

where the $TPR = \frac{\text{True Positive}}{\text{True Positive+False Negative}}$, and the $FPR = \frac{\text{False Positive}}{\text{False Positive+True Negative}}$.

Since treatment options for COVID-19 are limited and many drugs are still being investigated, we obtained the drugs that have entered clinical trial phase II or above from the ChEMBL database as the true positive set. To evaluate the discrimination power of the DTSEA method, we plotted a ROC curve based on the true positive set. This enables us to visualize the discrimination power of the DTSEA method.

2.3.3. The overall prediction consistency across multiple datasets

To validate our predictions, we employed the inter-rater concordance (or agreement) method. The method measures the overall consistency of predictions by assessing the agreement between multiple rows of data [33]. Initially, the inter-rater concordance is defined as the extent of agreement among multiple raters when rating the responses of the same cohort. Here, we shall consider the DTSEA per sample as a rating witness or a scoring machine (akin to the "judge" in the original concept) while the drugs as candidates. As a result, we adapted the inter-rater concordance method to validate the prediction performance of DTSEA. In this study, Kendall's W and Cronbach's alpha were utilized as independent statistics to determine the level of concordance. A high score indicates a strong level of agreement between multiple predictions for the same drugs. Conversely, a low score indicates that the predictions are inconsistent or heterogeneous.

The first statistic utilized in this study was Kendall's coefficient of concordance W, which ranges from 0 (no agreement) to 1 (complete agreement). We assumed that there were n objects (drugs) and m judges (datasets predicted by DTSEA), which formed a matrix (M) with n rows and m columns. The elements of matrix M were the NESs calculated by DTSEA. The W is defined as:

$$W = \frac{12\sum_{i=1}^{n}(R_i - M)^2}{m^2(n^3 - n)} \tag{4}$$

where the R_i denotes the sum of the rank r_i for each row i, and the mean value of total ranks R is $M = \frac{1}{n}\sum_{i=1}^{n} R_i$.

In contrast, Kendall's W measures the consistency of the rank or the relative position across multiple predictions. Nevertheless, Kendall's W does not provide a complete picture of the variance and error [34]. To complement the results obtained from Kendall's W, we then employ Cronbach's α to measure the agreement between multiple judges (datasets predicted by DTSEA). Cronbach's α is a popular measurement in social sciences [35] and clinical research [36]. The α is defined as [37]:

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum_{j=1}^{k}\sigma_j^2}{\sigma_{\bar{X}}^2}\right) \tag{5}$$

where k denotes the number of columns (judges) in the scoring matrix M, $\sigma_{\bar{X}}^2$ denotes the variance of the sum of each row (each drug), $\sum_{j=1}^{k}\sigma_j^2$ denotes the sum of the variances of each column j.

3. Results

3.1. Identification of COVID-19 candidate drugs based on DTSEA

We first applied DTSEA to COVID-19 gene expression datasets obtained from GSE183071 [23], and the entire results were provided in Supplementary Table S3. Among the top 50 most significant drugs identified, we observed that 24 drugs were being investigated (including

through case report (CR), silicon prediction (SILC), review (RV) in the literature, vitro experiment (VITRO), and vivo experiment (VIVO)), while 18 drugs were being evaluated in clinical trials (phases I-IV (P1–P4)) (Fig. 2A). Through comprehensive analysis, only 16% of these drugs have not been previously studied in the context of COVID-19 (Fig. 2B). These results indicate that our method could effectively recall COVID-19-related drugs.

Next, we focused on the top ten most significant drugs (Table 1). Interestingly, nine out of ten drugs have been studied experimentally, with three of them being evaluated in clinical trials. In particular, Ataluren, a COVID-19 antagonist with the highest potency, received conditional FDA approval in 2014 for treating Duchenne muscular dystrophy and cystic fibrosis [38]. The drug works by inhibiting the release factor complex (RFC) termination activity, acting at or before the hydrolysis step of RNA strands [39]. Despite the previous findings indicating that RFC termination was not an effective target against SARS-CoV-2, Ataluren showed a 46% inhibition in the cell experiment [40]. The associations between these drugs and COVID-19 are presented in Supplementary Table S4.

Since the platform GPL30569 only sequences immune-related genes and the conductors of GSE183071 focused on immune response [23], the expression dataset is relatively small and unable to fully explain the effect of the drug visually. Therefore, we adopted GSE164805, a genome-wide sequencing dataset that covers all drug targets to demonstrate the inhibition effects of Ataluren. We drew an expression heatmap in Fig. 3A, with the samples in GSE164805 as columns and the Ataluren-related genes (its targets and neighbors in the gene functional network) as rows. The heatmap clearly shows clusters of expression. Therefore, we divided the clusters into four categories and performed enrichment analyses for each category using the pathway data provided by KEGG.

The Ataluren-related genes are significantly enriched in several pathways. The effectiveness of Ataluren is supported by the enrichment of three out of four clusters (clusters #1, #3, and #4) in the Coronavirus disease — COVID-19 pathway, which suggests that the Ataluren-related genes directly impact the disease. Other enriched pathways, such as ribosome pathway (hsa03010; cluster #1), Notch signaling pathway (hsa04330; cluster #2), MAPK signaling pathway (hsa04010; cluster #2), and nucleotide excision repair pathway (hsa03420; cluster #3), were essential to the human immune response, which plays a central role in the inflammatory response [42–45]. Hence, the first candidate drug, Ataluren, may be a potential treatment for COVID-19.

The second potent COVID-19 drug, Carfilzomib, is a highly effective treatment for adults with relapsed or refractory multiple myeloma [46]. By inhibiting the chymotrypsin-like catalytic protease, Carfilzomib shows efficiently reduces cellular proliferation. Some evidence suggests that Carfilzomib is one of the potential inhibitors of SARS-CoV-2 main protease [47]. The Carfilzomib-related genes are significantly enriched in several immunological reaction pathways and virus reaction pathways (shown in 3B), including the homologous recombination pathway (hsa03440; cluster #1), the antigen processing and presentation pathway (hsa04612; cluster #2), and the Huntington's disease pathway (hsa05016; cluster #4).

The third potent COVID-19 drug, Aripiprazole, is an atypical treatment for various mood and psychotic disorders, such as schizophrenia, bipolar disorder, major depressive disorder, and agitation. Aripiprazole stabilizes the levels of neurotransmitters to improve psychotic symptoms and has a high affinity for and acts as a partial agonist for at least one serotonin and dopamine receptor [48]. Recent studies have reported the role of the serotonin family in the immune response to specific viral infections [49–51], including human immunodeficiency virus (HIV), reovirus, and chikungunya virus. Consistent with these studies, Aripiprazole-related genes are significantly enriched at the Dopaminergic synapse pathway (hsa04728; cluster #2) and several immunological pathways, including the chemokine signaling pathway (hsa04062; clusters #2, #3, and #4), and the Calcium signaling pathway (hsa04020; clusters #1, #2, and #4) (Fig. 3C). Therefore, these drugs could be considered potential candidates for further investigations into their efficiency in treating COVID-19.

3.2. Evaluation of the performance of DTSEA

3.2.1. Evaluating the network proximity between drug targets and disease genes

In the gene functional network, the DTSEA assumes that the most effective drugs are those with a closer affinity to the specified disease genes. Based on this assumption, the DTSEA suggests that a drug with a large ES should have a shorter distance between the drug targets and disease genes in the gene functional interaction network.

The first criterion we used was the shortest path length among the drug-disease module pairs. Based on the results in the previous subsection, we manually created six artificial groups to compare the mean distance between the targets of each drug in a specific group and disease genes. The six categories were named alphabetically, which correspond to the top 50 drugs (Group A), the top 51 to 200 drugs (Group B), the top 201 to 400 drugs (Group C), the positive ES drugs with a not significant p-value ($p > .015$ and $p < 0.40$; Group D), the positive ES drugs with a nearly random p-value ($p > 0.50$; Group E), and the negative ES drugs (Group F), respectively. We calculated the average distances and assumed the variances were unequal, so we performed Welch's one-way ANOVA test.

It revealed that (Fig. 4A) the main effect was statistically significant and substantial across the drug groups ($F = 413.4$, $p = 5.79e-101$, $\eta^2 = 0.56$). Then, we used the post hoc Games-Howell analysis to

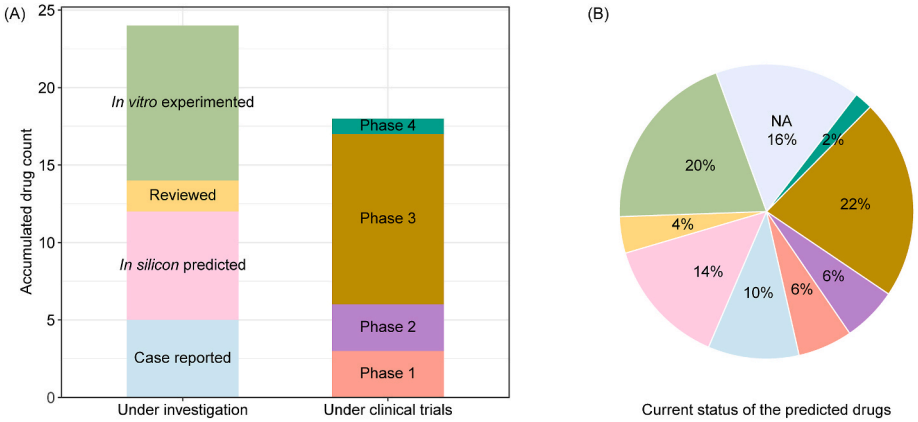

Fig. 2. Summary of the top 50 predicted drugs. (A) Histogram shows the drug counts of current status of the top 50 predicted drugs. **(B)** Pie chart shows the percentages of current status of the top 50 predicted drugs.

Table 1
List of the top 10 compounds and their indications in the prediction set using GSE183071.

DrugBank ID	Name	Main indications	Evidence ID	LOE	NES	FDR
DB05016	Ataluren	Duchenne muscular dystrophy	34904435*	VITRO	3.208	8.715e-25
DB08889	Carfilzomib	Multiple myeloma	32315171*	SILC	3.113	2.011e-19
DB00188	Bortezomib	Multiple myeloma	33551422*	CR	3.113	2.011e-19
DB09570	Ixazomib	Neoplasms	35409348*	VITRO	3.113	2.011e-19
DB11991	Oprozomib	Neoplasms			3.113	2.011e-19
DB12010	Fostamatinib	Hemorrhage; thrombocytopenia	NCT04579393#	P2	2.800	2.868e-28
DB00201	Caffeine	ADHD; pain; apnea; migraine disorder	33094705*	SILC	2.772	7.068e-11
DB00806	Pentoxifylline	Cardiovascular disease	NCT04433988#	P1	2.761	7.068e-11
DB01238	Aripiprazole	Major depressive disorder; autistic disorder; bipolar disorder; schizophrenia	33990069*	CR	2.684	6.213e-10
DB01017	Minocycline	Infections; periodontitis; rosacea; conjunctivitis; acne vulgaris; psittacosis; fever	NCT05077813#	P2	2.671	1.094e-09

Note: In the table, the levels of evidence (LOE) rating for the investigation process among the predicted drugs is based on the following criterions: case report (CR), silicon prediction (SILC), review (RV) in literatures, vitro experiment (VITRO), vivo experiment (VIVO), and clinical trial phases I-IV (P1–P4), which incorporates the multiple perspectives illustrated by Burns et al. [41]. Furthermore, if the presented drug is under clinical trials, then the maximum phase the drug has reached will be shown here. The column Evidence ID represents the reference to the LOE. In this column, if the corresponding drug is under the clinical trial phases, then the registered IDs on clinical trials were provided; otherwise, PubMed IDs were provided.
Note *: PubMed ID.
Note #: Clinical trail ID.

compare the average distances between each pair of drug groups, the results showed that each pair of groups differed significantly at $p < 0.001$ except for the group pairs of Group A-B ($p = .978$), Group A-C ($p = 0.162$), and Group B–C ($p = 0.309$). These results suggest that the top drugs predicted by DTSEA are close to COVID-19 related genes in the network. Notably, the trend analysis in Fig. 4A demonstrated that the linear model was significant at the monomial level ($p = 7.14e-131$) and the quadratic level ($p = 7.47e-26$) with a substantial explanatory power of $R^2 = 0.56$, but nonsignificant at the cubic level ($p = 0.542$), indicating that the distance increases as the ranks of the results predicted by DTSEA get further backward.

Next, we quantified the network proximity between drug targets and disease genes in the network by using the network separation metric (mentioned in Methods), which is the main idea of the complementary exposure component analysis [52]. Specification, for each drug in each group (A-F), we respectively calculated network separation metric between drug targets gene set (called drug module) and disease gene set (called disease module) in the network. Through comparing the network separation metric across all the drug groups, the one-way ANOVA showed a significant but small effect ($F = 36.4$, $p = 1.92e-30$, $\eta^2 = 0.07$, Fig. 4B). Moreover, the post hoc Games-Howell test indicated that each pair of drug groups differed significantly at $p < 0.001$ except for the group pairs of Group C-D ($p = 1.00$), Group C-E ($p = 0.18$), and Group E-F ($p = 0.30$), demonstrating that DTSEA predicted drugs have a close network-based distance with COVID-19 related genes. We further performed a trend analysis on the linear model and observed that the model was significant at the monomial level ($p = 1.90e-20$), the quadratic level ($p = 0.011$), and the cubic level ($p = 0.026$), indicating a positive correlation between the prediction order and the separation metric. Consistent with the original hypothesis of the DTSEA, the above results indicate that DTSEA is reliable and meets the topological assumptions.

3.2.2. Evaluating the robustness of DTSEA

As mentioned above, the converging evidence strongly supports the effectiveness of the DTSEA in network topology. However, the question remains as to whether DTSEA can maintain its performance when facing uncertainty and implausible alternative hypotheses [53]. To evaluate the robustness of the method, we conducted simulations involving five types of node failures and five types of edge failures with 50 repetitions under each condition. In each simulation, we randomly eliminated a certain percentage of nodes or edges.

The simulation experiment revealed that edge and node deletion were effective methods for assessing robustness, and were relatively simple to execute. Thus, we generated uncertainties by deleting random edges and nodes. After constructing the uncertainty models, we applied DTSEA to the pruned networks. We then compared the top 50 and 100 predicted drug lists generated from the pruned graphs with the unpruned graph. As shown in Fig. 4C, the percentage of overlapped drugs is relatively stable regardless of the amount of pruning. Even if we removed half of the edges or nodes in the network, nearly half of the original predictions were preserved, indicating the high robustness of DTSEA.

3.2.3. Evaluating the reproducibility and accuracy of DTSEA

Several studies have identified heterogeneous genes associated with COVID-19 [54]. One possible explanation for this heterogeneity is those innate and adaptive immune responses to viruses caused by human genetic variants [55]. Given the individual differences mentioned above, it is essential to evaluate the reproducibility of DTSEA by comparing it with other COVID-19 datasets. We shall define one disease dataset as a pipeline. Until now, we have presented the prediction results in a single pipeline. The following validation is performed to assess the reproducibility of DTSEA across different pipelines.

We applied DTSEA to five datasets (pipelines), four of which were expression profiles obtained from the GEO database. The other was a COVID-19-related gene list collected from NCBI (described in Methods). Initially, we identified the top 50 drugs predicted by DTSEA from each dataset and then intersected these drugs to obtain a set of 19 drugs that appeared in at least three of the result sets (see Supplementary Table S5). As shown in Fig. 5A and 5B, approximately 79% (15/19) of these drugs were either under investigation or in clinical trials for COVID-19. Furthermore, the Venn diagram in Fig. 5C indicates that the predictions were reasonably consistent.

We then evaluated the consistency of predictions for each pair of five pipelines as described in Methods section. As shown in Fig. 5D, the overall consistency was high ($\alpha = 0.877$, $W = 0.662$), indicating a strong interdependence among the result pipelines. Furthermore, we observed significant moderate to high pairwise correlations. Then we evaluated and compared the prediction performance of DTSEA across five different pipelines by analyzing ROC curve. The AUC values for all five pipelines exceeded 0.85 (Fig. 5E), indicating that there is a greater than 85% likelihood of identifying a potential drug-disease association using DTSEA. The results of the five different pipelines provide insight into the reliability of DTSEA across various pipelines.

3.3. Extended analysis for single target drugs

Due to the high sensitivity and reliability of DTSEA, we obtained promising prediction results in the preceding steps. However, when only one gene was available for certain drugs in step 3 of the DTSEA workflow (left part of Fig. 1C), the enrichment result produced by GSEA was

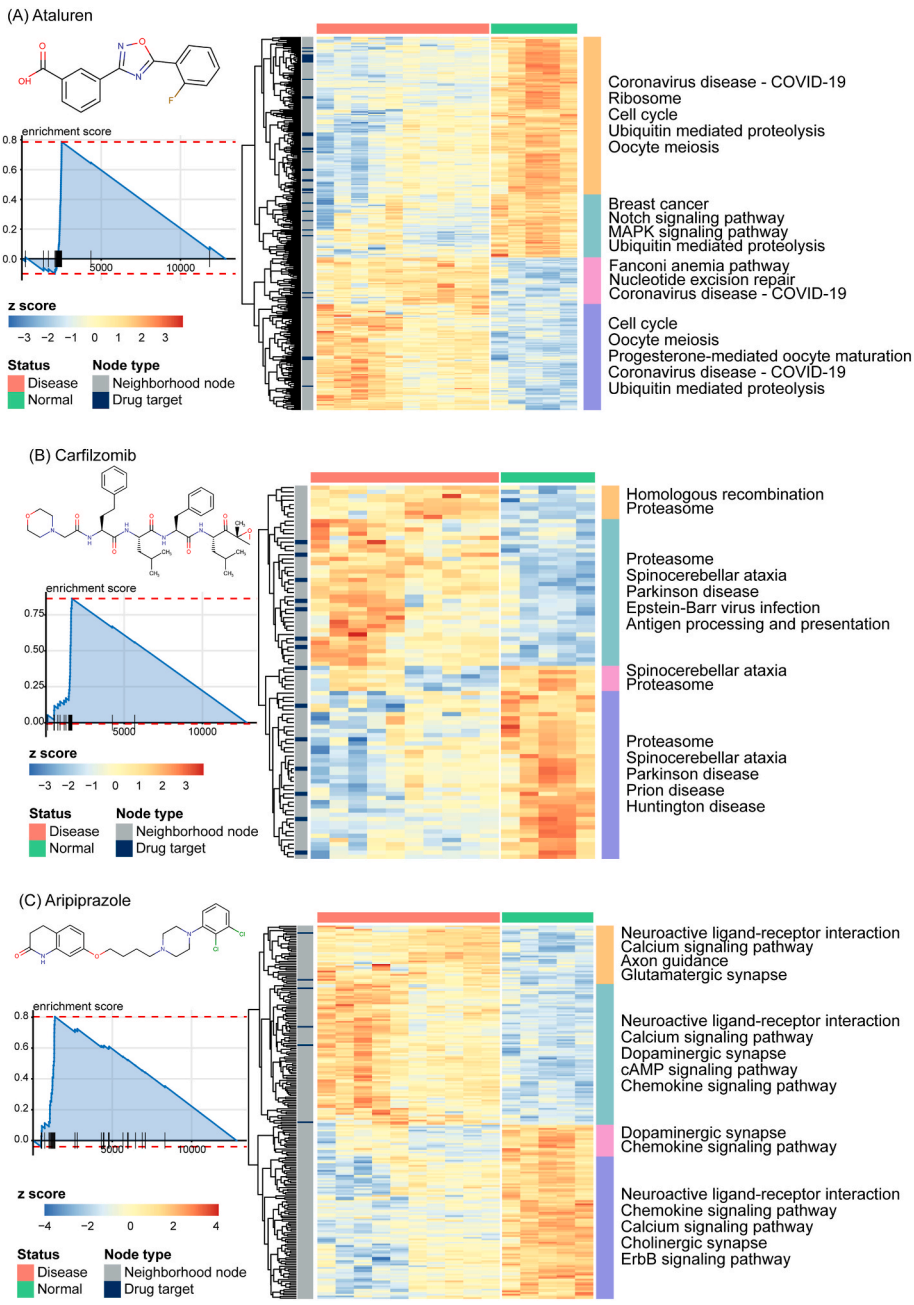

Fig. 3. Visualization of Ataluren, Carfilzomib, and Aripiprazole. (A) The left half plots the drug chemical structure and drug target gene set enrichment analysis of Ataluren. The right half plots the heatmap of Ataluren-related genes (its targets and neighbors in the gene functional network). The genes were clustered into four categories, and the genes in each category were annotated into the KEGG pathways. **(B)** The left half plots the drug chemical structure and drug target gene set enrichment analysis of Carfilzomib. The right half plots the heatmap of Carfilzomib-related genes (its targets and neighbors in the gene functional network). The genes were clustered into four categories, and the genes in each category were annotated into the KEGG pathways. **(C)** The left half plots the drug chemical structure and drug target gene set enrichment analysis of Aripiprazole. The right half plots the heatmap of Aripiprazole-related genes (its targets and neighbors in the gene functional network). The genes were clustered into four categories, and the genes in each category were annotated into the KEGG pathways.

unreliable. The GSEA achieves maximum power by analyzing a set of genes (gene set) rather than a single gene. Thus, no reliable result is obtained if only one gene is present in the gene set [28,56]. As a result, we only retained multi-target drugs, while eliminating single-target drugs.

We observed that over half of the drugs in our dataset had only one target (see Fig. 6A). As a result, we manually extended our local drug target database using this portion of data. However, we only validated the prediction results for a limited number of drugs in this part. These drugs may receive approval documents, emergency use authorizations, or results in phase 3 clinical trials. Firstly, we created a list of candidates to be evaluated using the thumb rule (see Supplementary Table S6). We then collected the predicted target set for each drug from ChEMBL and conducted DTSEA on these drugs. We obtained positive results for all drugs except Propofol (Fig. 6B).

3.4. Extended analysis for breast cancer

In the previous section, we utilized multiple validation methods to assess the effectiveness of DTSEA in specific COVID-19 sets. We then evaluated whether DTSEA could identify candidate drugs for other diseases to assess its generalization. Specifically, we applied DTSEA to six breast cancer datasets, comprising four DEG sets identified from GEO breast cancer gene expression datasets (ID: GSE20711, GSE15852, GSE42568, and GSE10780), as well as two other breast cancer-related gene sets acquired from the Online Mendelian Inheritance in Man (OMIM) database and the Candidate Cancer Gene Database (CCGD).

We first assessed the correlation coefficient between prediction results among different pipelines to validate the consistency and validity of predictions. We observed a high level of consistency in these predictions among six datasets (Fig. 7A, $W = 0.625$, Cronbach's $\alpha = 0.847$), with significant moderate to high pairwise correlations.

We then assessed the validity of the prediction results using ROC

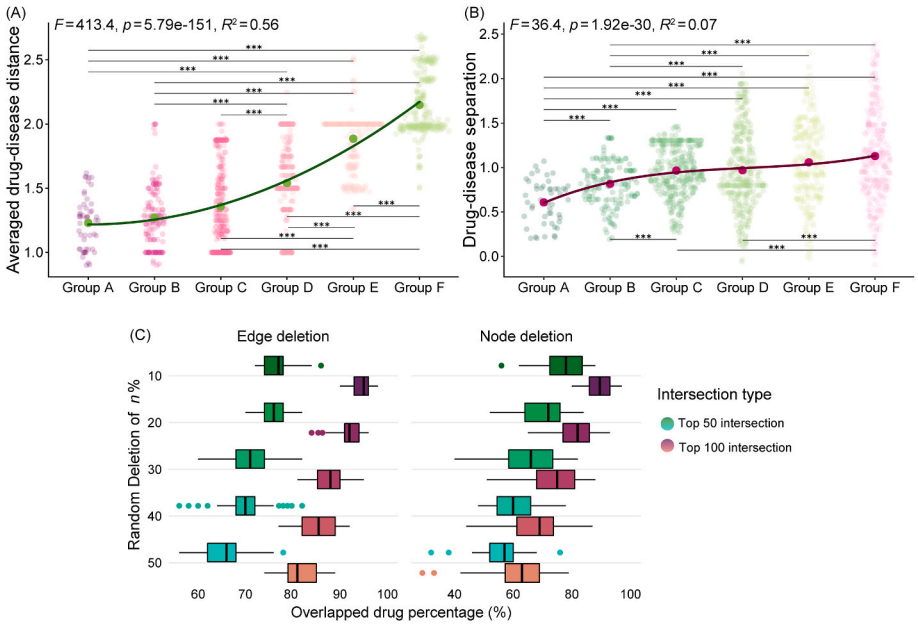

Fig. 4. Validations of DTSEA. Comparisons of the (**A**) averaged distance and (**B**) network separation between the targets of predicted drugs and disease genes. The Welch's ANOVA and the trend analysis were applied to compare marginal mean difference among groups of drugs. The predicted drugs were categorized into six drug groups named in alphabetical order and corresponded to: the top 50 drugs (Group A), the top 51 to 200 drugs (Group B), the top 201 to 400 drugs (Group C), the positive ES drugs with a not good p-value ($p > 0.15$ and $p < 0.40$; Group D), the positive ES drugs with a nearly random p-value ($p > 0.50$; Group E), and the negative ES drugs (Group F), respectively. **(C)** Randomly deleted edges and nodes revealed the robustness of the DTSEA. In the combined box plot, the left part shows the intersection percentage of multiple scenarios of edge deletion, while the right part shows the percentage of random deletions. In each subplot, we performed two intersections with the top 50 and 100 sets. To illustrate the relationships clearly, we manually shifted several pixels in each group of dots.

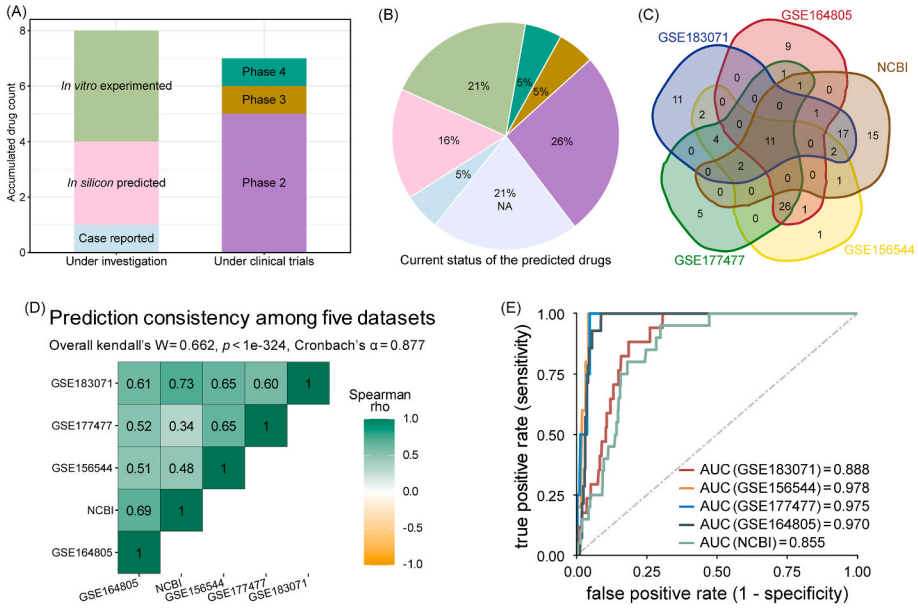

Fig. 5. Reproducibility of DTSEA in five datasets (pipelines). (A) Histogram shows the summary of the overlapped drugs predicted by at least three COVID-19 related datasets. **(B)** Pie chart shows the percentages of current status of the overlapped drugs. **(C)** Venn diagram shows the count of overlapping drugs across five COVID-19 related datasets. **(D)** Diagonal matrix shows the Spearman correlation coefficient (rho) between the drug lists predicted by each pair of datasets. **(E)** ROC curves show the predicted power of DTSEA across five datasets.

curve. Consistent with our previous analyses, we used drugs approved for breast cancer by the FDA or in Phase II clinical trials as the criterion. Our analysis of the ROC curves and AUC values (Fig. 7B) confirmed the validity of DTSEA for breast cancer. These findings support DTSEA for drug repurposing in other diseases.

4. Discussion

This study proposed the DTSEA method as an effective approach to repurpose existing drugs for screening potential indications. We introduced a novel metric for drug-disease distance and expanded the usage of the GSEA method. As it stands, the DTSEA can provide unique insight into finding candidate drugs for human diseases.

Recent evidence suggested that Cilgavimab and Bamlanivimab could be potential monoclonal antibody treatments for the two latest COVID-19 variants [57]. However, DTSEA cannot validate the two drugs as our combined drug target database contains only one target for each of them. A large drug target set is ideal for enrichment analysis rather than a small one, which is the limitation of the DTSEA method. Therefore, it would be ideal for the DTSEA to give unbiased results for drugs with a more comprehensive set of targets. On the one hand, the biased result is unavoidable for the DTSEA unless large drug target sets become more widely available. On the other hand, precise medicine does not entail obtaining an abundance of targets according to its propaganda [58].

The reliability of the latest evidence of some drugs is still uncertain. As an example of Remdesivir, the clinical trial conducted by Peking Union Medical College was statistically nonsignificant [59], whereas the NIH finding was significant [60]. Determining which drug is genuinely effective is challenging since the algorithm does not predict drug response in vivo or in vitro. As a compromise, we decided to keep all significant results regardless of the nonsignificant ones.

In addition, the experimental determination of drug-protein

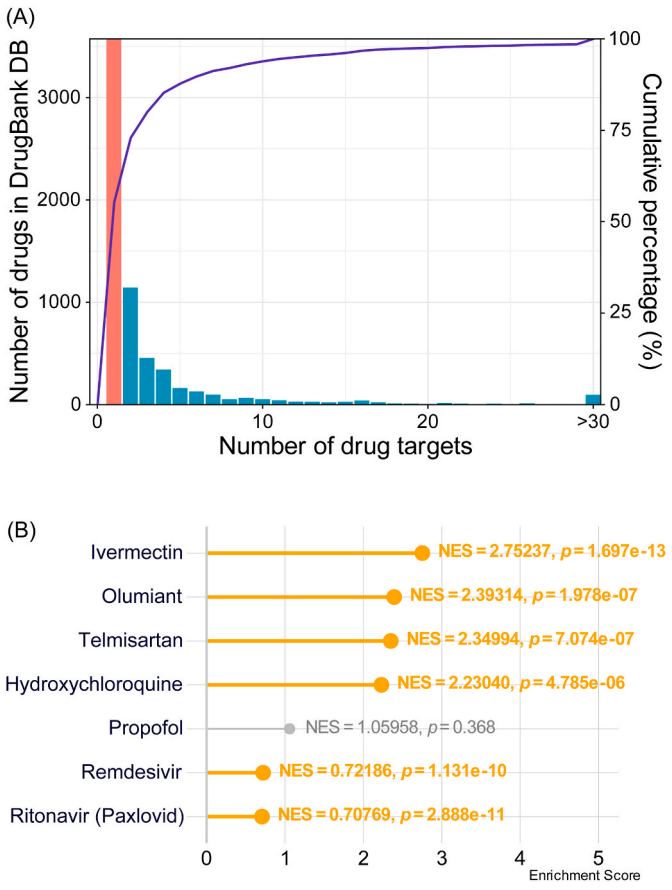

Fig. 6. Extended analysis for single target drugs. (A) The distribution of drug target of the preamble drug set. **(B)** Validation of several potent drugs. For each drug, the NES and p-value were shown next to the bar plot.

interactions is time-consuming, expensive, and limited to small-scale research [61–63], which plays a crucial and predetermined role in drug discovery [64]. As the downstream process of drug discovery, we should make the best use of all available data to improve predictions, given the scarcity of the data, rather than deleting over half of the information in the batch prediction process (Fig. 6A). Because of the deletions, the method is incomparable to other methods.

Currently, only a few drugs have been approved for COVID-19, and exploring the relationship between target genes and disease-related genes in the human interactome [65]. Despite this, few computational approaches and clinical trials were conducted during the pandemic [66], and these barriers will need to be removed in the future.

Data availability statement

The COVID-19 gene expression profiles were obtained from the GEO database (ID: GSE156544, GSE164805, GSE177477, and GSE183071), and the other one is a 190 COVID-19-related gene list obtained from NCBI in June 2022 (https://www.ncbi.nlm.nih.gov/gene/?term=coronavirus+related+%5Bproperties%5D). The 11 genes associated with both COVID-19 and three comorbidities are available in the publications cited in the manuscript.

The clinical trial data of drugs and drug indications were collected from the ChEMBL database (https://www.ebi.ac.uk/chembl/, Version CHEMBL30, March 2022). The drug-target interactions were downloaded and integrated from ChEMBL and DrugBank (https://go.drugbank.com/DrugBank, Version 5.1.9, Jan 2022) databases. The human pathways data were retrieved from seven databases, including KEGG, Reactome, Biocarta, NCI, SPIKE, HumanCyc, and Panther.

We wrapped the core function into an R package named DTSEA (Drug Target Set Enrichment Analysis), which is freely available under the GPL-v2 license.

Authors' contributions

Junwei Han and QK conceived and designed the study. YS, JW, and XL developed the method. JQ, BP, and Junling Huang analyzed the data and implemented the methodology. JL and XZ provided constructive discussions. Junwei Han and YS drafted the manuscript. All the authors read and agreed to the manuscript.

Funding

This work was supported in part by the National Natural Science Foundation of China (grant no. 62072145); and the Natural Science Foundation of Heilongjiang Province (grant no. LH2019C042).

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Declaration of competing interest

The authors declare no competing interests.

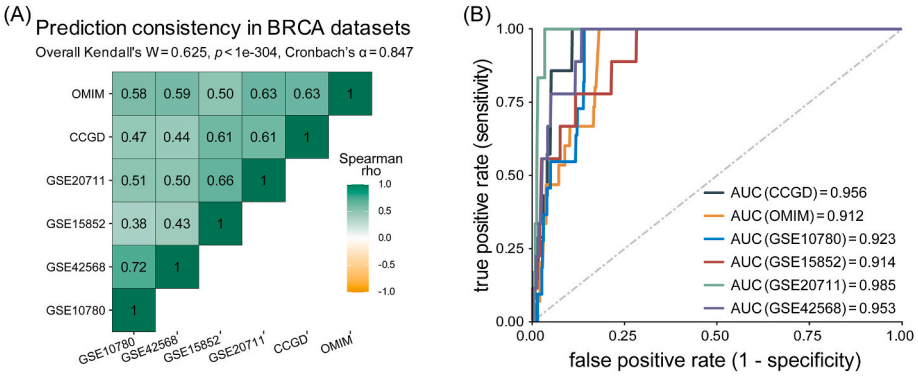

Fig. 7. Extended analysis for breast cancer. (A) Diagonal matrix shows the Spearman correlation coefficient (rho) between the drug lists predicted by each pair of datasets. **(B)** ROC curves show the predicted power of DTSEA across six breast cancer datasets.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2023.106969.

References

[1] M. Worobey, J.I. Levy, L. Malpica Serrano, A. Crits-Christoph, J.E. Pekar, S. A. Goldstein, A.L. Rasmussen, M.U.G. Kraemer, C. Newman, M.P.G. Koopmans, M. A. Suchard, J.O. Wertheim, P. Lemey, D.L. Robertson, R.F. Garry, E.C. Holmes, A. Rambaut, K.G. Andersen, The huanan seafood wholesale market in wuhan was the early epicenter of the COVID-19 pandemic, Science (80-) 377 (2022) 951–959, https://doi.org/10.1126/science.abp8715.

[2] D. Benvenuto, M. Giovanetti, A. Ciccozzi, S. Spoto, S. Angeletti, M. Ciccozzi, The 2019-new coronavirus epidemic: evidence for virus evolution, J. Med. Virol. 92 (2020) 455–459, https://doi.org/10.1002/jmv.25688.

[3] A.L. Rasmussen, On the origins of SARS-CoV-2, Nat. Med. 27 (2021) 9, https://doi.org/10.1038/s41591-020-01205-5.

[4] A. June, S. Adjei, K. Hong, N.M. Molinari, L. Bull-otterson, U.A. Ajani, A. V Gundlapalli, A.M. Harris, J. Hsu, S.S. Kadri, J. Starnes, K. Yeoman, T.K. Boehmer, Mortality risk among patients hospitalized primarily for COVID-19 during the Omicron and Delta variant pandemic periods — United States, MMWR Morb Mortal Wkly Rep 71 (2022) 1182–1189.

[5] E. Petersen, M. Koopmans, U. Go, D.H. Hamer, N. Petrosillo, F. Castelli, M. Storgaard, S. Al Khalili, L. Simonsen, Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics, Lancet Infect. Dis. 20 (2020), https://doi.org/10.1016/S1473-3099(20)30484-9 e238–e244.

[6] M. Rudrapal, S.J. Khairnar, A.G. Jadhav, Drug Repurposing (DR): An Emerging Approach in Drug Discovery, in: Drug Repurposing - Hypothesis, Mol. Asp. Ther. Appl, 1st ed., IntechOpen, London, UK, London, 2020, pp. 2–4, https://doi.org/10.5772/intechopen.93193.

[7] T. Cihlar, R.L. Mackman, Journey of remdesivir from the inhibition of hepatitis C virus to the treatment of COVID-19, Antivir. Ther. 27 (2022) 1–12, https://doi.org/10.1177/13596535221082773.

[8] T.K. Warren, R. Jordan, M.K. Lo, A.S. Ray, R.L. Mackman, V. Soloveva, D. Siegel, M. Perron, R. Bannister, H.C. Hui, N. Larson, R. Strickley, J. Wells, K.S. Stuthman, S.A. Van Tongeren, N.L. Garza, G. Donnelly, A.C. Shurtleff, C.J. Retterer, D. Gharaibeh, R. Zamani, T. Kenny, B.P. Eaton, E. Grimes, L.S. Welch, L. Gomba, C. L. Wilhelmsen, D.K. Nichols, J.E. Nuss, E.R. Nagle, J.R. Kugelman, G. Palacios, E. Doerffler, S. Neville, E. Carra, M.O. Clarke, L. Zhang, W. Lew, B. Ross, Q. Wang, K. Chun, L. Wolfe, D. Babusis, Y. Park, K.M. Stray, I. Trancheva, J.Y. Feng, O. Barauskas, Y. Xu, P. Wong, M.R. Braun, M. Flint, L.K. McMullan, S.-S. Chen, R. Fearns, S. Swaminathan, D.L. Mayers, C.F. Spiropoulou, W.A. Lee, S.T. Nichol, T. Cihlar, S. Bavari, Therapeutic efficacy of the small molecule GS-5734 against Ebola virus in rhesus monkeys, Nature 531 (2016) 381–385, https://doi.org/10.1038/nature17180.

[9] Y.N. Lamb, Remdesivir: first approval, Drugs 80 (2020) 1355–1363, https://doi.org/10.1007/s40265-020-01378-w.

[10] A. Fernández-Torras, M. Duran-Frigola, P. Aloy, Encircling the regions of the pharmacogenomic landscape that determine drug response, Genome Med. 11 (2019) 1–15, https://doi.org/10.1186/s13073-019-0626-x.

[11] Z. Liu, H. Fang, K. Reagan, X. Xu, D.L. Mendrick, W. Slikker, W. Tong, In silico drug repositioning-what we need to know, Drug Discov. Today 18 (2013) 110–115, https://doi.org/10.1016/j.drudis.2012.08.005.

[12] D.M. Gysi, Í. do Valle, M. Zitnik, A. Ameli, X. Gan, O. Varol, S.D. Ghiassian, J. J. Patten, R.A. Davey, J. Loscalzo, A.-L. Barabási, Network medicine framework for identifying drug-repurposing opportunities for COVID-19, Proc. Natl. Acad. Sci. USA 118 (2021), e2025581118, https://doi.org/10.1073/pnas.2025581118.

[13] U.S. Saini, E.E. Papalexakis, A peek into the hidden layers of a convolutional neural network through a factorization lens, ArXiv Prepr. (2018). http://arxiv.org/abs/1806.02012.

[14] Y. Abo-zeid, N.S. Ismail, G.R. McLean, N.M. Hamdy, A molecular docking study repurposes FDA approved iron oxide nanoparticles to treat and control COVID-19 infection, Eur. J. Pharmaceut. Sci. 153 (2020), 105465, https://doi.org/10.1016/j.ejps.2020.105465.

[15] M.R. Senger, T.C.S. Evangelista, R.F. Dantas, M.V. da S. Santana, L.C.S. Gonçalves, L.R. de S. Neto, S.B. Ferreira, F.P. Silva-Junior, COVID-19: molecular targets, drug repurposing and new avenues for drug discovery. https://doi.org/10.1590/0074-02760200254, 2020.

[16] X. Jalencas, J. Mestres, Identification of similar binding sites to detect distant polypharmacology, Mol. Inform. 32 (2013) 976–990, https://doi.org/10.1002/minf.201300082.

[17] L. Rodrigues, R.B. Cunha, T. Vassilevskaia, M. Viveiros, C. Cunha, Drug repurposing for COVID-19: a review and a novel strategy to identify new targets and potential drug candidates, Molecules 27 (2022), https://doi.org/10.3390/molecules27092723.

[18] Y. Masoudi-Sobhanzadeh, H. Esmaeili, A. Masoudi-Nejad, A fuzzy logic-based computational method for the repurposing of drugs against COVID-19, Bioimpacts 12 (2022) 315–324, https://doi.org/10.34172/bi.2021.40.

[19] J. Wu, X. Li, Q. Wang, J. Han, DRviaSPCN: a software package for drug repurposing in cancer via a subpathway crosstalk network, Bioinformatics 38 (2022) 4975–4977, https://doi.org/10.1093/bioinformatics/btac611.

[20] X. Han, Q. Kong, C. Liu, L. Cheng, J. Han, SubtypeDrug: a software package for prioritization of candidate cancer subtype-specific drugs, Bioinformatics (2021), https://doi.org/10.1093/bioinformatics/btab011.

[21] J. Di, B. Zheng, Q. Kong, Y. Jiang, S. Liu, Y. Yang, X. Han, Y. Sheng, Y. Zhang, L. Cheng, J. Han, Prioritization of candidate cancer drugs based on a drug functional similarity network constructed by integrating pathway activities and drug activities, Mol. Oncol. 13 (2019) 2259–2277, https://doi.org/10.1002/1878-0261.12564.

[22] J. Han, X. Han, Q. Kong, L. Cheng, psSubpathway: a software package for flexible identification of phenotype-specific subpathways in cancer progression, Bioinformatics 36 (2020) 2303–2305, https://doi.org/10.1093/bioinformatics/btz894.

[23] A. Gómez-Carballa, I. Rivero-Calle, J. Pardo-Seco, J. Gómez-Rial, C. Rivero-Velasco, N. Rodríguez-Núñez, G. Barbeito-Castiñeiras, H. Pérez-Freixo, M. Cebey-López, R. Barral-Arca, C. Rodriguez-Tenreiro, A. Dacosta-Urbieta, X. Bello, S. Pischedda, M.J. Currás-Tuala, S. Viz-Lasheras, F. Martinón-Torres, A. Salas, GEN-COVID study group, A multi-tissue study of immune gene expression profiling highlights the key role of the nasal epithelium in COVID-19 severity, Environ. Res. 210 (2022) 112890, https://doi.org/10.1016/j.envres.2022.112890.

[24] S. Feng, F. Song, W. Guo, J. Tan, X. Zhang, F. Qiao, J. Guo, L. Zhang, X. Jia, Potential genes associated with COVID-19 and comorbidity, Int. J. Med. Sci. 19 (2022) 402–415, https://doi.org/10.7150/ijms.67815.

[25] A. Gaulton, A. Hersey, M.L. Nowotka, A. Patricia Bento, J. Chambers, D. Mendez, P. Mutowo, F. Atkinson, L.J. Bellis, E. Cibrian-Uhalte, M. Davies, N. Dedman, A. Karlsson, M.P. Magarinos, J.P. Overington, G. Papadatos, I. Smit, A.R. Leach, The ChEMBL database in 2017, Nucleic Acids Res. 45 (2017) D945–D954, https://doi.org/10.1093/nar/gkw1074.

[26] D.S. Wishart, Y.D. Feunang, A.C. Guo, E.J. Lo, A. Marcu, J.R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda, N. Assempour, I. Iynkkaran, Y. Liu, A. MacIejewski, N. Gale, A. Wilson, L. Chin, R. Cummings, Di Le, A. Pon, C. Knox, M. Wilson, DrugBank 5.0: a major update to the DrugBank database for 2018, Nucleic Acids Res. 46 (2018) D1074–D1082, https://doi.org/10.1093/nar/gkx1037.

[27] S. Köhler, S. Bauer, D. Horn, P.N. Robinson, Walking the interactome for prioritization of candidate disease genes, Am. J. Hum. Genet. 82 (2008) 949–958, https://doi.org/10.1016/j.ajhg.2008.02.013.

[28] A. Subramanian, P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, M.A. Gillette, A. Paulovich, S.L. Pomeroy, T.R. Golub, E.S. Lander, J.P. Mesirov, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, Proc. Natl. Acad. Sci. U.S.A. 102 (2005) 15545–15550, https://doi.org/10.1073/pnas.0506580102.

[29] Y. Benjamini, D. Yekutieli, The control of the false discovery rate in multiple testing under dependency, Ann. Stat. 29 (2001) 1165–1188, https://doi.org/10.1214/aos/1013699998.

[30] G. Korotkevich, V. Sukhov, N. Budin, B. Shpak, M.N. Arttyomov, A. Sergushichev, Fast Gene Set Enrichment Analysis, BioRxiv Prepr, 2016, pp. 1–29, https://doi.org/10.1101/060012.

[31] J. Menche, A. Sharma, M. Kitsak, S.D. Ghiassian, M. Vidal, J. Loscalzo, A. L. Barabási, Uncovering disease-disease relationships through the incomplete interactome, Science (80-) 347 (2015) 841, https://doi.org/10.1126/science.1257601.

[32] T. Fawcett, An introduction to ROC analysis, Pattern Recogn. Lett. 27 (2006) 861–874, https://doi.org/10.1016/j.patrec.2005.10.010.

[33] L.D. Wahyuni, G. Gumela, H. Maulana, Interrater Reliability: comparison of essay's tests and scoring rubrics, J. Phys. Conf. Ser. 1933 (2021), https://doi.org/10.1088/1742-6596/1933/1/012081.

[34] M. Tavakol, R. Dennick, Making sense of Cronbach's alpha, Int. J. Med. Educ. 2 (2011) 53–55, https://doi.org/10.5116/ijme.4dfb.8dfd.

[35] D.G. Bonett, T.A. Wright, Cronbach's alpha reliability: interval estimation, hypothesis testing, and sample size planning, J. Organ. Behav. 36 (2014) 3–15, https://doi.org/10.1002/job.1960.

[36] M. Heo, N. Kim, M.S. Faith, Statistical power as a function of Cronbach alpha of instrument questionnaire items, BMC Med. Res. Methodol. 15 (2015) 1–9, https://doi.org/10.1186/s12874-015-0070-6.

[37] E. Cho, Making reliability reliable: a systematic approach to reliability coefficients, Organ. Res. Methods 19 (2016) 651–682, https://doi.org/10.1177/1094428116656239.

[38] N.J. Ryan, Ataluren: first global approval, Drugs 74 (2014) 1709–1714, https://doi.org/10.1007/s40265-014-0287-4.

[39] S. Huang, A. Bhattacharya, M.D. Ghelfi, H. Li, C. Fritsch, D.M. Chenoweth, Y. E. Goldman, B.S. Cooperman, Ataluren binds to multiple protein synthesis apparatus sites and competitively inhibits release factor-dependent termination, Nat. Commun. 13 (2022) 1–14, https://doi.org/10.1038/s41467-022-30080-6.

[40] M. Dasovich, J. Zhuo, J.A. Goodman, A. Thomas, R.L. McPherson, A.K. Jayabalan, V.F. Busa, S.J. Cheng, B.A. Murphy, K.R. Redinger, Y.M.O. Alhammad, A.R. Fehr, T. Tsukamoto, B.S. Slusher, J. Bosch, H. Wei, A.K.L. Leung, High-throughput activity assay for screening inhibitors of the SARS-CoV-2 Mac1 macrodomain, ACS Chem. Biol. 17 (2022) 17–23, https://doi.org/10.1021/acschembio.1c00721.

[41] P.B. Burns, R.J. Rohrich, K.C. Chung, The levels of evidence and their role in evidence-based medicine, Plast. Reconstr. Surg. 128 (2011) 305–310, https://doi.org/10.1097/PRS.0b013e318219c171.

[42] L. Almeida, A. Dhillon-LaBrooy, C.N. Castro, N. Adossa, G.M. Carriche, M. Guderian, S. Lippens, S. Dennerlein, C. Hesse, B.N. Lambrecht, L. Berod, L. Schauser, B.R. Blazar, M. Kalesse, R. Müller, L.F. Moita, T. Sparwasser, Ribosome-targeting antibiotics impair T cell effector function and ameliorate autoimmunity by blocking mitochondrial protein synthesis, Immunity 54 (2021) 68–83.e6, https://doi.org/10.1016/j.immuni.2020.11.001.

[43] T. Ikawa, H. Kawamoto, A.W. Goldrath, C. Murre, E proteins and Notch signaling cooperate to promote T cell lineage specification and commitment, J. Exp. Med. 203 (2006) 1329–1342, https://doi.org/10.1084/jem.20060268.
[44] M. Rincón, R.A. Flavell, R.A. Davis, The JNK and p38 MAP kinase signaling pathways in T cell-mediated immune responses, Free Radic. Biol. Med. 28 (2000) 1328–1337, https://doi.org/10.1016/S0891-5849(00)00219-7.
[45] Y. He, N. Pasupala, H. Zhi, B. Dorjbal, I. Hussain, H.M. Shih, S. Bhattacharyya, R. Biswas, M. Miljkovic, O.J. Semmes, T.A. Waldmann, A.L. Snow, C.Z. Giam, NF-κB–induced R-loop accumulation and DNA damage select for nucleotide excision repair deficiencies in adult T cell leukemia, Proc. Natl. Acad. Sci. U.S.A. 118 (2021) 1–9, https://doi.org/10.1073/pnas.2005568118.
[46] K.M. Kortuem, A.K. Stewart, Carfilzomib, Blood 121 (2013) 893–897, https://doi.org/10.1182/blood-2012-10-459883.
[47] J. Wang, Fast identification of possible drug treatment of coronavirus disease-19 (COVID-19) through computational drug repurposing study, J. Chem. Inf. Model. 60 (2020) 3277–3286, https://doi.org/10.1021/acs.jcim.0c00179.
[48] H.A. Nasrallah, Atypical antipsychotic-induced metabolic side effects: insights from receptor-binding profiles, Mol. Psychiatr. 13 (2008) 27–35, https://doi.org/10.1038/sj.mp.4002066.
[49] Y. Pashaei, Drug repurposing of selective serotonin reuptake inhibitors: could these drugs help fight COVID-19 and save lives? J. Clin. Neurosci. 88 (2021) 163–172, https://doi.org/10.1016/j.jocn.2021.03.010.
[50] N. Herr, C. Bode, D. Duerschmied, The effects of serotonin in immune cells, Front. Cardiovasc. Med. 4 (2017) 1–11, https://doi.org/10.3389/fcvm.2017.00048.
[51] B.A. Mainou, A.W. Ashbrook, E.C. Smith, D.C. Dorset, M.R. Denison, T.S. Dermody, Serotonin receptor agonist 5-nonyloxytryptamine alters the kinetics of reovirus cell entry, J. Virol. 89 (2015) 8701–8712, https://doi.org/10.1128/jvi.00739-15.
[52] F. Cheng, I.A. Kovács, A.L. Barabási, Network-based prediction of drug combinations, Nat. Commun. 10 (2019), https://doi.org/10.1038/s41467-019-09186-x.
[53] J. Stelling, U. Sauer, Z. Szallasi, F.J. Doyle, J. Doyle, Cell 118 (2004) 675–685, https://doi.org/10.1016/j.cell.2004.09.008.
[54] I. Fricke-Galindo, R. Falfán-Valencia, Genetics insight for COVID-19 susceptibility and severity: a review, Front. Immunol. 12 (2021) 1–11, https://doi.org/10.3389/fimmu.2021.622176.
[55] I.G. Ovsyannikova, I.H. Haralambieva, S.N. Crooke, G.A. Poland, R.B. Kennedy, The role of host genetics in the immune response to SARS-CoV-2 and COVID-19 susceptibility and severity, Immunol. Rev. 296 (2020) 205–219, https://doi.org/10.1111/imr.12897.
[56] J. Shi, M. Walker, Gene set enrichment analysis (GSEA) for interpreting gene expression profiles, Curr. Bioinf. 2 (2008) 133–137, https://doi.org/10.2174/157489307780618231.
[57] D.J. Sheward, C. Kim, J. Fischbach, S. Muschiol, R.A. Ehling, N.K. Björkström, G.B. K. Hedestam, S.T. Reddy, J. Albert, T.P. Peacock, B. Murrell, Evasion of neutralising antibodies by Omicron sublineage BA.2.75, Lancet Infect. Dis. (2022), https://doi.org/10.1016/S1473-3099(22)00524-2.
[58] A.A. Alzu'bi, L. Zhou, V.J.M. Watzlaf, Genetic variations and precision medicine, Perspect. Heal. Inf. Manag. 16 (2019).
[59] Y. Wang, D. Zhang, G. Du, R. Du, J. Zhao, Y. Jin, S. Fu, L. Gao, Z. Cheng, Q. Lu, Y. Hu, G. Luo, K. Wang, Y. Lu, H. Li, S. Wang, S. Ruan, C. Yang, C. Mei, Y. Wang, D. Ding, F. Wu, X. Tang, X. Ye, Y. Ye, B. Liu, J. Yang, W. Yin, A. Wang, G. Fan, F. Zhou, Z. Liu, X. Gu, J. Xu, L. Shang, Y. Zhang, L. Cao, T. Guo, Y. Wan, H. Qin, Y. Jiang, T. Jaki, F.G. Hayden, P.W. Horby, B. Cao, C. Wang, Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multicentre trial, Lancet 395 (2020) 1569–1578, https://doi.org/10.1016/S0140-6736(20)31022-9.
[60] J.H. Beigel, K.M. Tomashek, L.E. Dodd, A.K. Mehta, B.S. Zingman, A.C. Kalil, E. Hohmann, H.Y. Chu, A. Luetkemeyer, S. Kline, D. Lopez de Castilla, R. W. Finberg, K. Dierberg, V. Tapson, L. Hsieh, T.F. Patterson, R. Paredes, D. A. Sweeney, W.R. Short, G. Touloumi, D.C. Lye, N. Ohmagari, M. Oh, G.M. Ruiz-Palacios, T. Benfield, G. Fätkenheuer, M.G. Kortepeter, R.L. Atmar, C.B. Creech, J. Lundgren, A.G. Babiker, S. Pett, J.D. Neaton, T.H. Burgess, T. Bonnett, M. Green, M. Makowski, A. Osinusi, S. Nayak, H.C. Lane, Remdesivir for the treatment of covid-19 — final report, N. Engl. J. Med. 383 (2020) 1813–1826, https://doi.org/10.1056/nejmoa2007764.
[61] X. Chen, C.C. Yan, X. Zhang, X. Zhang, F. Dai, J. Yin, Y. Zhang, Drug-target interaction prediction: databases, web servers and computational models, Briefings Bioinf. 17 (2016) 696–712, https://doi.org/10.1093/bib/bbv066.
[62] R. Jafari, H. Almqvist, H. Axelsson, M. Ignatushchenko, T. Lundbäck, P. Nordlund, D.M. Molina, The cellular thermal shift assay for evaluating drug target interactions in cells, Nat. Protoc. 9 (2014) 2100–2122, https://doi.org/10.1038/nprot.2014.138.
[63] L. Wang, Z.H. You, L.P. Li, X. Yan, W. Zhang, Incorporating chemical sub-structures and protein evolutionary information for inferring drug-target interactions, Sci. Rep. 10 (2020) 1–11, https://doi.org/10.1038/s41598-020-62891-2.
[64] R. Chen, X. Liu, S. Jin, J. Lin, J. Liu, Machine learning for drug-target interaction prediction, Molecules 23 (2018) 1–15, https://doi.org/10.3390/molecules23092208.
[65] A.S. Imami, R.E. McCullumsmith, S.M. O'Donovan, Strategies to identify candidate repurposable drugs: COVID-19 treatment as a case example, Transl. Psychiatry 11 (2021), https://doi.org/10.1038/s41398-021-01724-w.
[66] G. Galindez, J. Matschinske, T.D. Rose, S. Sadegh, M. Salgado-Albarrán, J. Späth, J. Baumbach, J.K. Pauling, Lessons from the COVID-19 pandemic for advancing computational drug repurposing strategies, Nat. Comput. Sci. 1 (2021) 33–41, https://doi.org/10.1038/s43588-020-00007-6.