


RESEARCH

Open Access



# Development of complemented comprehensive networks for rapid screening of repurposable drugs applicable to new emerging disease outbreaks

Yonghyun Nam<sup>1</sup>, Anastasia Lucas<sup>2</sup>, Jae-Seung Yun<sup>1,3</sup>, Seung Mi Lee<sup>1,4</sup>, Ji Won Park<sup>1,5</sup>, Ziqi Chen<sup>6</sup>, Brian Lee<sup>1</sup>, Xia Ning<sup>6,8,9</sup>, Li Shen<sup>1,10</sup>, Anurag Verma<sup>7\*</sup> and Dokyoon Kim<sup>1,10\*</sup> 

## Abstract

**Background** Computational drug repurposing is crucial for identifying candidate therapeutic medications to address the urgent need for developing treatments for newly emerging infectious diseases. The recent COVID-19 pandemic has taught us the importance of rapidly discovering candidate drugs and providing them to medical and pharmaceutical experts for further investigation. Network-based approaches can provide repurposable drugs quickly by leveraging comprehensive relationships among biological components. However, in a case of newly emerging disease, applying a repurposing methods with only pre-existing knowledge networks may prove inadequate due to the insufficiency of information flow caused by the novel nature of the disease.

**Methods** We proposed a network-based complementary linkage method for drug repurposing to solve the lack of incoming new disease-specific information in knowledge networks. We simulate our method under the controlled repurposing scenario that we faced in the early stage of the COVID-19 pandemic. First, the disease-gene-drug multi-layered network was constructed as the backbone network by fusing comprehensive knowledge database. Then, complementary information for COVID-19, containing data on 18 comorbid diseases and 17 relevant proteins, was collected from publications or preprint servers as of May 2020. We estimated connections between the novel COVID-19 node and the backbone network to construct a complemented network. Network-based drug scoring for COVID-19 was performed by applying graph-based semi-supervised learning, and the resulting scores were used to validate prioritized drugs for population-scale electronic health records-based medication analyses.

**Results** The backbone networks consisted of 591 diseases, 26,681 proteins, and 2,173 drug nodes based on pre-pandemic knowledge. After incorporating the 35 entities comprised of complemented information into the backbone network, drug scoring screened top 30 potential repurposable drugs for COVID-19. The prioritized drugs were subsequently analyzed in electronic health records obtained from patients in the Penn Medicine COVID-19 Registry as of October 2021 and 8 of these were found to be statistically associated with a COVID-19 phenotype.

\*Correspondence:

Anurag Verma

Anurag.Verma@pennmedicine.upenn.edu

Dokyoon Kim

dokyoon.kim@pennmedicine.upenn.edu

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



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

**Conclusion** We found that 8 of the 30 drugs identified by graph-based scoring on complemented networks as potential candidates for COVID-19 repurposing were additionally supported by real-world patient data in follow-up analyses. These results show that our network-based complementary linkage method and drug scoring algorithm are promising strategies for identifying candidate repurposable drugs when new emerging disease outbreaks.

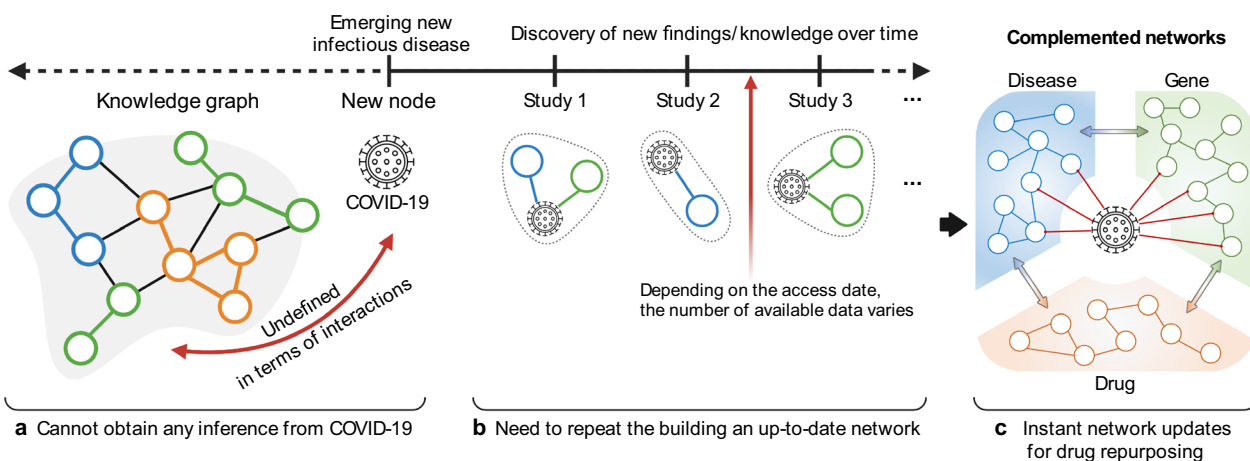
**Keywords** Drug repurposing, Network medicine, Graph-based semi-supervised learning, COVID-19

**Introduction**

During the recent COVID-19 pandemic caused by the severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), many researchers and pharmaceutical companies have conducted numerous studies for developing treatments and vaccines. In particular, many researchers attempt to repurpose known drugs to treat patients with SARS-CoV-2 infection because drug repurposing is cheaper and quicker than conventional drug discovery. Drug repurposing aims to find new indicators in already-approved drugs that could be used for other diseases [1–3]. Several drugs have been successfully repositioned for COVID-19, such as remdesivir (initially developed to treat the Ebola virus) and dexamethasone (used in conditions for anti-inflammatory and immunosuppressant effects) [4–6]. Although the global COVID-19 pandemic is gradually shifting to an endemic stage thanks to the development of vaccines and treatments, the spread of coronavirus is still ongoing as of September 2022. One of the lessons we have learned from the recent pandemic is that it is important to rapidly discover a list of candidate drugs and provide it to experts in the medical or pharmaceutical field can investigate the potential of the candidate drugs for use in new indications [7].

As knowledge of biological mechanisms advances and biomedical knowledge is well collected, more accurate and precise computational drug repurposing based on well-curated data has become possible [8]. One computational repurposing framework is a network-based approach that can recommend candidate drugs by observing the complex relationships among biological entities such as drugs, genes, and diseases. Their complex and heterogeneous interactions can be represented by topological structures among nodes and edges in a graph [9–11].

Even though biomedical/pharmaceutical data sources are more readily available than ever, what if a new infectious disease emerges and there is no information about the new disease in the previously accumulated database? This is the scenario the scientific and medical communities faced in the early stages of the COVID-19 pandemic (urgent need for discovering therapeutic treatment for COVID-19). In a network-based approach, the new disease node, i.e. COVID-19, could be introduced in an existing biological network to apply network-based drug repurposing. However, if the new disease is disconnected with the existing network, meaning COVID-19 is undefined in terms of interactions with other components, the disconnected node cannot provide sufficient pharmaceutical



**Fig. 1** Limitations of network-based approaches in the face of the emerging new infectious disease. **a** New disease (COVID-19) cannot be connected with the knowledge network using data collected before emerging COVID-19. **b** Although the novel associations related to COVID-19 can discover by researchers over time, the final network for drug repurposing can differ depending on the selected dataset. **c** We developed a new method of updating the network instantly with the complementary dataset (discovered findings from studies) to find candidate repurposable drugs for COVID-19

evidence or inferences (Fig. 1a). As COVID-19-related research progressed gradually, relational information (such as pathogenesis of COVID-19 and related target genes/proteins) were discovered over time (Fig. 1b). To connect the COVID-19 node to the previously constructed network, the network needs to be rebuilt or updated by incorporating the newly found information. As the information related to COVID-19 discovered, various networks including latest information can be built or updated depending on which association information is used (Fig. 1b). However, it is inefficient to repeat the entire process of building an up-to-date network every time new data sources come in when much new information is revealed and reported from the researchers continuously. Obviously, a more sophisticated network will help predict repurposable drugs with more therapeutic potential [12, 13]. However, in a public health emergency as we have experienced, it is also crucial to provide even potential evidences for candidate drugs to allow pharmaceutical experts to conduct early-stage trials, even if the candidate drugs were predicted from a relatively less elaborate networks. Inspired by this, we propose a network-based method for rapid screening of repurposable drugs that enables to efficiently incorporate the complementary information for a new entity into existing networks (Fig. 1c).

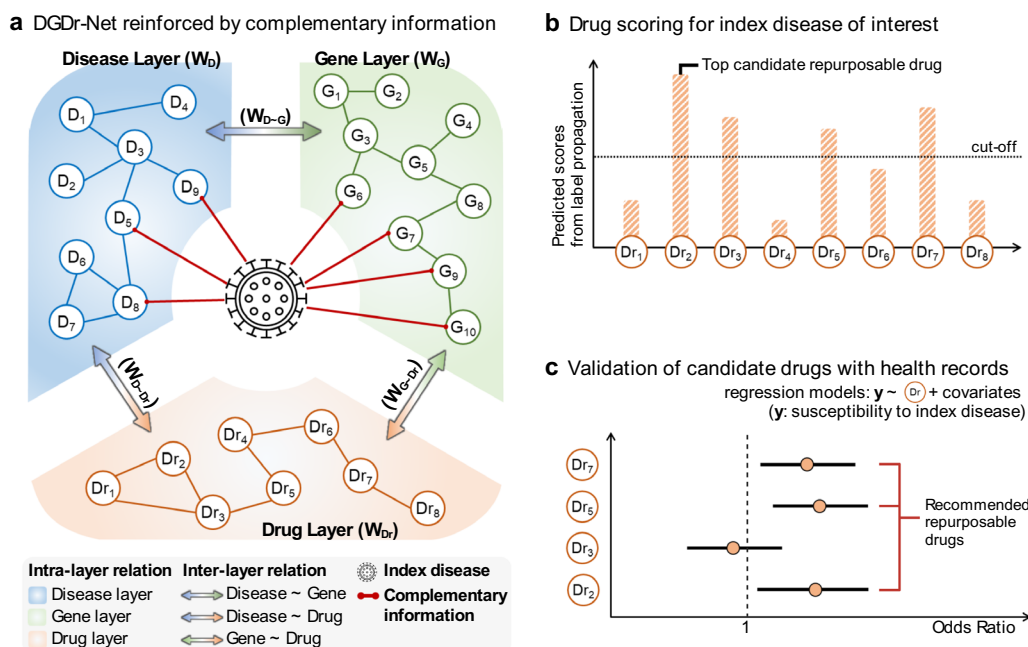
In this study, we assumed that we faced the initial/early stage of the COVID-19 pandemic. We collected the relationship information for COVID-19 reported as of May 2020. Through this paper, we showed the simulation of how to overcome the lack of information related to new infectious diseases and how to validate the candidate repurposable drugs from the proposed complementary network. First, a backbone network without COVID-19 related information was created. The backbone network consisted of heterogeneous components such as diseases, genes, and drugs which connected with each other based on calculated proximity. The backbone network was processed as a multi-layered network that consists of three different single networks including a disease-disease network, a protein-protein interaction network and a drug-drug network. To introduce the novel disease into the constructed backbone network, a network-based complementary linkage method was developed to estimate the auxiliary connections between the new disease node and the heterogeneous multi-layered network. In our previous study, we already developed complementary method for enhancing a single disease network or a single drug network to improve their connectivity, but there was a limitation in the previous method of not being able to make estimating multiple connections at once—the previous method can only connect one edge per iteration, and thus the overall quality of estimated complementary edges is dependent on the order of connected edges [14, 15]. We

improved the complementary linkage method to enable estimating a batch of multiple connections at once by applying the enhanced multi-layered network with heterogeneous or homogenous data. Next, with the complemented network complemented by the novel disease node and its estimated connections, repurposable drug were screened via graph-based semi-supervised learning, which propagates label information along with the multi-layered topological structure. The label propagation algorithm can produce a ranked list of prioritized candidate drugs with normalized scores. We then took the candidate drugs with the highest scores and looked for evidence of associations with patient medication orders and COVID-19 related phenotypes using electronic health record information from the Penn Medicine health system.

## Methods

### Overview of proposed network-based drug repurposing method

We propose a network-based drug repurposing method for rapid screening to respond to the situation of an emerging new disease. The proposed method can quickly update and augment a knowledge-driven comprehensive multi-layered network with novel disease relationship data and then prioritize the candidate repurposable drugs based on the complemented network. First, we constructed the multi-layered network as a backbone by collecting data from publicly available databases (Additional file 2: Table. S1). As described above, we assumed that COVID-19 emerged as a new disease node that was not connected to the multi-layered backbone network. Additional relational information about COVID-19 to be augmented to the backbone network was collected from papers published or shared on preprint servers (medRxiv and bioRxiv) before May 2020 and contains diseases comorbid with and genes relevant to COVID-19 [16]. A network-based complementary linkage method was developed to estimate connections between the novel COVID-19 disease node and the backbone network (Fig. 2a). The complementary linkage method can determine whether the additional relational information can harm the backbone network when the arbitrary edges provided by users are connected. Then, graph-based semi-supervised learning (SSL) was applied to prioritize the repurposable candidate drugs for COVID-19 (Fig. 2b). The graph-based SSL can predict candidate drugs by leveraging the underlying structure of complemented network when only one label information was given [17–19]. The prioritized list of candidate drugs was validated with electronic health records at Penn Medicine.

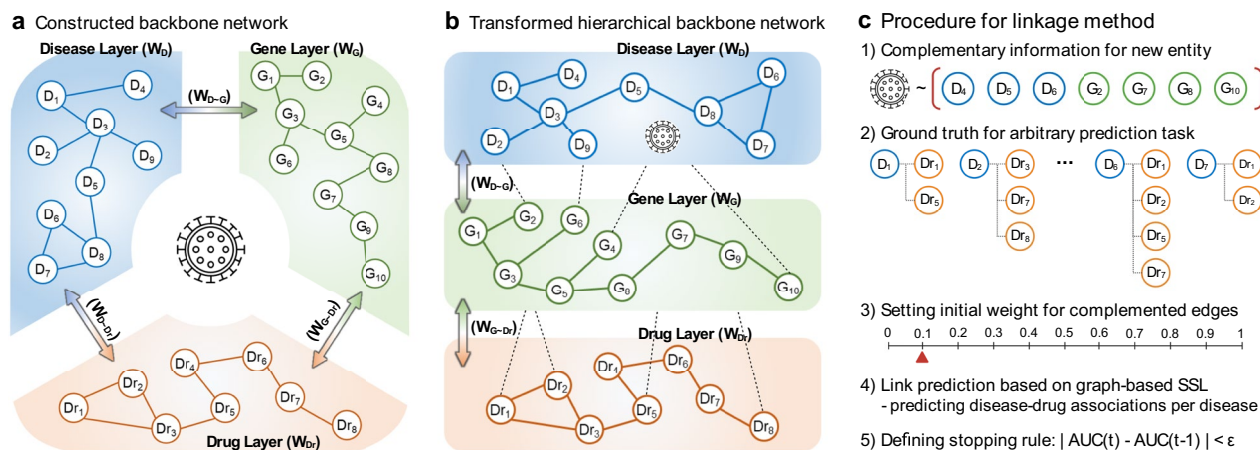


**Fig. 2** Schematic description of proposed network-based drug repurposing method. **a** Complementing the backbone network with COVID-19 information. **b** Drug scores for COVID-19 by applying graph-based semi-supervised learning. **c** Association analysis with the predicted candidate drugs and electronic health records to evaluate drug scoring

### Construction of backbone network

The Disease-Gene-Drug Network was constructed as a backbone network to represent relationships between different biological components. The backbone network is a multi-layered heterogeneous graph,  $G = (V, W, S)$ , where the set of nodes  $V$  represents diseases, genes, and drugs according to the set of layers  $S = \{D, G, Dr\}$  respectively, and the similarity matrix  $W$  represents the relationships within and across layers (Fig. 3a). Since the

network is multi-layered, we defined intra-layer relations and inter-layer relations by decomposing the similarity matrix  $W$  as  $W^{(intra)}$  and  $W^{(inter)}$ . The intra-layer relation depicts a single network such as a disease-disease network, a protein-protein interaction network, or a drug-drug network. The similarity for single network was quantified by calculating the cosine similarity using the respective association vectors. For example, similarity between diseases were calculated by disease-gene



**Fig. 3** Systematic process of complementary linkage method for multi-layered network

association vectors. The inter-layer relation represents the connections between different single networks (different layers), which consist of disease-gene association, disease-drug association, and drug-gene associations [20]. More details about constructing network are described in Additional file 1: Extended method, and Additional file 2: Table S1.

**Network-wise complementary linkage method for multi-layered network**

The network-based complementary linkage method can estimate connections between a single disease node of COVID-19 and the backbone network. The strategies of the proposed linkage method are as follows: (a) adding initial information provided by users for estimating auxiliary connections (estimated edges), (b) prediction tasks are defined within the backbone network to learn how to estimate new edges between backbone network and novel node, (c) the properties of backbone network are defined as a loss function, and (d) user-provided auxiliary connections are allowed in a complementary process, provided it does not compromise the pre-defined properties of the backbone network (stopping rule).

The proposed steps were used to augment the novel COVID-19 node into the backbone network. (a) 18 comorbid diseases and 17 related genes were collected from the literature and used initial information (Additional file 2: Table S2). (b) The prediction task was defined as predicting disease-drug associations in the backbone network. The original tripartite multi-layered backbone network was transformed into the hierarchical layered network to facilitate these disease-drug association predictions, ordered by disease, gene, and drug layer (Fig. 3b). When transforming into a hierarchical network, connections between disease and drug layers were deleted and used as the ground truth for the prediction task during complementary process. Notably, we predict the list of drugs when a single disease node is given as a label. The number of iterations is the same as the number of nodes in the disease-disease network. (c) The measuring property of the transformed backbone network was defined as the area under the receiver operating characteristic (AUC) for an index disease of interest. (d) 35 user-provided auxiliary connections were allowed when the average AUC was not decreased. The connection strength increased the search range by 0.1 units (Fig. 3c).

**Drug scoring algorithm with complemented network**

A graph-based SSL was employed to prioritize candidate repurposable drugs with the complemented network. Semi-supervised approaches can be used even if the label information is insufficient compared to the conventional

supervised approaches that always require a lot of label information [19]. The scoring algorithm propagates the given label information to the underlying structure of the complemented network. It is a more suitable case for employing graph-based SSL when a new disease such as COVID-19 in the previously assumed situation has no therapeutic agent.

The formulation of the scoring algorithm is as follows. Consider we have  $m$  diseases,  $n$  proteins, and  $k$  drugs in complemented network,  $G = (V, W)$ , with set of nodes  $V (= V_D \cup V_G \cup V_{Dr})$  corresponding to the  $|V| (= m + n + k)$  nodes. Let  $y = (y_1, \dots, y_{|V|})^T$  denote the initial label set of nodes, and  $f = [f^D, f^G, f^{Dr}]^T = (f_1^D, \dots, f_m^D, f_1^G, \dots, f_n^G, f_1^{Dr}, \dots, f_k^{Dr})^T$  denote the set of resulting scores. Unlike the general classification problem where the target variable has a binary label ('+1' or '-1'), the problem setting of scoring in a semi-supervised approach has a unary label ('+1') only. More specifically, a disease node of COVID-19 ( $v_{COVID}^D$ ) is set to a unary label  $y_{COVID} \in \{+1\}$ , and the other nodes set to zero ( $y \setminus y_{COVID} \in \{0\}$ ). In graph-based SSL, there are two assumptions: (a) a loss function that predicted scores in unlabeled nodes should be close to the given label of  $y_i$  in labeled nodes and (b) a smoothness condition that predicted scores in adjacent unlabeled nodes should be close to each other. These assumptions are reflected by the quadratic objective function in Eq. (1) where the graph Laplacian  $L$  is defined as  $L = D - W$ ,  $D = \text{diag}(\sum_j w_{ij})$  is diagonal degree matrix of  $W$ , and the user-specified parameter  $\mu$  trades off loss and smoothness (to reduce computational complexity,  $\mu$  is set to  $1/\|L\|_1$  in this study).

$$\min (f - y)^T (f - y) + \mu f^T L f \tag{1}$$

By minimizing objective function in Eq. (1), the closed-form solution becomes  $f = (I + \frac{L}{\|L\|_1})^{-1} y$ . The predicted score  $f$  is produced for all three single layers, but since we are only interested in repurposable drugs in the drug layer, the value of  $f^{Dr}$  is transformed to range from 0 to 1 as  $\hat{f}^{Dr} = \frac{f^{Dr} - \min(f^{Dr})}{\max(f^{Dr}) - \min(f^{Dr})}$ . All drugs are sorted in descending order according to the transformed scores  $\hat{f}^{Dr}$  to prioritize the repurposable drugs for the index disease of interest [15, 17].

**Validation of prediction results with candidate repurposable drugs**

The predicted results from the scoring algorithm can provide the prioritized scores of each drug, but it is hard to validate the potential of each candidate drug since

there is no pharmacological evidence related to the novel disease COVID-19. Although the listed candidate drugs for the COVID-19 have already been approved to treat other illnesses, it is unreasonable to conclude predict these drugs' efficacy to treat COVID-19 as these drugs have not yet undergone clinical trials for the COVID-19. However, during global health emergencies such as the COVID-19 pandemic, there is not enough time for clinical trials so an effective indirect verification method is needed. Therefore, candidate drugs were statistically validated using electronic health record (EHR) data.

First, we extracted the clinical records from the Penn Medicine COVID-19 Registry and assigned case/control phenotypes for COVID-19 susceptibility, hospitalization, and severity based on the COVID-19 Host Genetics Initiative phenotype definitions [21]. Then, we built logistic regression models for the effects of the candidate medications from the complemented network on phenotypes ( $y$ ) related to the index disease using each medication ( $\hat{f}^{Dr}$ ) as a predictor and adjusting for age, gender, and self-reported race as covariates:  $y_{\text{phenotype}} \sim \hat{f}_{[i]}^{Dr} + \text{age} + \text{gender} + \text{race}$ , where phenotype was one of the EHR-derived COVID-19 phenotypes (i.e., susceptibility, hospitalization, severity, and mortality), and  $\hat{f}_{[i]}^{Dr}$  is the  $[i]^{\text{th}}$  ranked candidate drug obtained from the complemented network.

The full study protocol for Penn Medicine EHR analysis was approved by the University of Pennsylvania Institutional Review Board (IRB) under the protocol for the study titled "Clinical, social, and genetic risk stratification for COVID-19 outcomes" (Protocol #844,360).

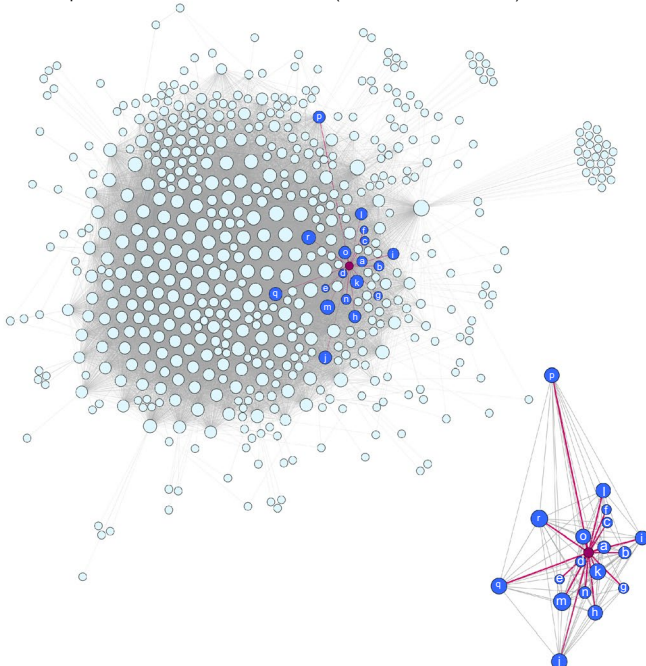
## Results

### Complementing backbone Disease-Gene-Drug network with COVID-19 information

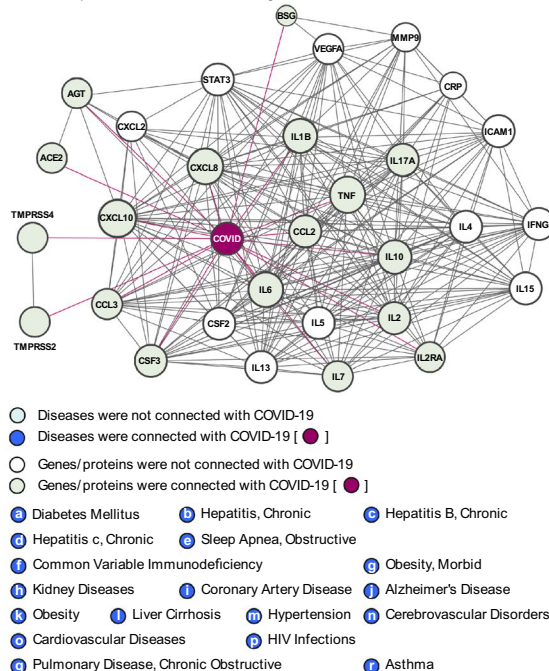
The backbone network was constructed with 591 diseases, 26,681 genes (proteins), and 2,173 drugs by collecting a list of components and relational data from public databases including the Comparative Toxicogenomics Database (CTD), the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING), and DrugBank [22–24]. The backbone disease-gene-drug network had three different single networks: disease-disease network, protein–protein interaction network, and drug-drug network. These single networks were then connected based on the relational data. Intra-layer relationships contained 22,855 disease-disease associations among 591 diseases, 841,068 interactions among 26,681 proteins, and 577,040 drug-drug associations among 2,173 drugs. There were 31,991 disease-gene associations, 76,889 disease-drug associations, 9,540 drug-gene associations in inter-layer relationships.

In order to incorporate the COVID-19 disease node into the backbone network, we collected

**a** Complemented disease-disease network (521 disease + COVID-19)



**b** Complemented COVID-19 and gene associations



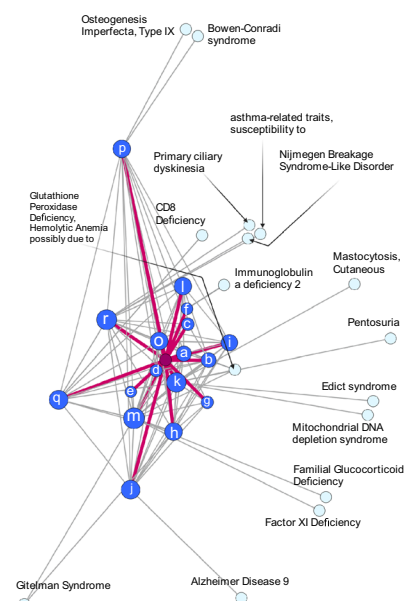
**Fig. 4** Complementing network with COVID-19 information: **a** complemented disease-disease network, **b** complemented protein–protein interaction network with COVID-19 disease node

initial information for auxiliary connections from the literatures: 18 diseases had been reported as comorbid with COVID-19 (e.g., diseases related to chronic hepatitis, HIV infections, obstructive sleep apnea, hypertension, obesity, diabetes mellitus, kidney diseases, common variable immunodeficiency, liver cirrhosis, coronary artery disease, chronic obstructive pulmonary disease, Alzheimer’s disease, asthma, cardiovascular disease, and cerebrovascular disorders) [16, 25, 26]. Also, we found 17 related genes for a biomarker of therapeutic evidence reported as of May 2020 (e.g., CCL2, TNF, IL10, CXCL8, IL6, IL1B, AGT, IL2, CXCL10, CCL3, Tmprss2, IL7, IL2RA, CSF3, Tmprss4, ACE2, and BSG). The connection strength (edge weights) for complementing edges was initially taken at a constant value of 0.1, and 0.1 units increased until the overall AUC decreased. In this analysis, the 35 complemented connections between COVID-19 and other nodes were estimated. Finally, the complemented disease disease-gene-drug network had 1,440,998 (=overall edges in backbone network+35 complemented edges) associations among 29,446 nodes. Figure 4 shows the complemented disease-gene-drug network with COVID-19. For easier visualization, we decomposed the entire network: the complemented disease-disease network (Fig. 3a) and the subset of the protein–protein interaction network (Fig. 4b). Note that there are no actual and complemented connections between COVID-19 and any drugs due to the defined problematic situations in this study.

**Internal quality check of the complemented network**

Before applying the drug scoring algorithm to prioritize candidate repurposable drugs, we conducted internal validation of the estimated connection between COVID-19 and other diseases. Although the reliability of the complemented network was verified by complementary process, we performed further quality check of connectivity between COVID-19 and other diseases/genes. To validate the connections focusing on COVID-19, a connectivity check within a single network was performed. First, to detect the community connected with COVID-19 node, the Louvain method for community detection was applied in the complemented single disease-disease network [27]. 33 diseases were belonging to same cluster with COVID-19. Notably, 18 directly- connected comorbidities (used for seed initial information) were included and the remaining 15 diseases were two-hop neighbors with COVID-19 in the cluster (Fig. 5a). Next, quality check for multi-layered network was performed. Scoring algorithms was applied to disease-gene complemented network for predicting COVID-19 related genes. Similar to method in drug scoring, the initial label was set to  $y_{COVID} \in \{+1\}$  and the others were set to  $\{0\}$  and applied graph-based SSL. We substituted the gene scores  $f^G$  from the entire predicted results  $f = [f^D, f^G]^T$ . The value of  $f^G$  is transformed to range from 0 to 1 via  $\hat{f}^G = \frac{f^G - \min(f^G)}{\max(f^G) - \min(f^G)}$ . All genes are sorted in descend-

**a** Member nodes in cluster with COVID-19



**b** 10 most relevant pathways sorted by p-value

Pathway name	Entities				Reactions	
	found	ratio	p-value	FDR	found	ratio
Interleukin-10 signaling	26 / 86	0.006	1.11E-16	4.33E-15	15 / 15	0.001
Interleukin-4 and Interleukin-13 signaling	24 / 211	0.014	1.11E-16	4.33E-15	39 / 47	0.003
Signaling by Interleukins	42 / 658	0.043	1.11E-16	4.33E-15	204 / 505	0.036
Cytokine Signaling in immune system	42 / 1107	0.073	1.11E-16	4.33E-15	248 / 726	0.052
Immune system	45 / 2698	0.178	1.11E-16	4.33E-15	255 / 1645	0.117
Chemokine receptors bind chemokines	9 / 57	0.004	1.34E-12	4.41E-11	4 / 19	0.001
RUNX1 and FOXP3 control the development of regulatory T lymphocytes (Tregs)	6 / 17	0.001	7.59E-11	2.21E-09	9 / 20	0.001
Peptide ligand-binding receptors	10 / 203	0.013	5.30E-09	1.27E-07	7 / 83	0.006
Interleukin-2 family signaling	6 / 47	0.003	3.10E-08	6.83E-07	42 / 59	0.004
Senescence-Associated Secretory Phenotype (SASP)	6 / 91	0.006	1.44E-06	2.73E-05	8 / 22	0.002

**Fig. 5** Internal quality check for complemented network: **a** Sub disease-disease network in community with COVID-19, **b** Top 10 significantly enriched pathways with selected 30 gene sets from scoring results

ing order according to the transformed scores  $\hat{f}^G$ . The overall gene scores were reported in Additional file 3. The top 30 of 26,681 genes sorted by gene score were TMPRSS4, TMPRSS2, ACE2, BSG, IL7, CSF3, CCL3, IL2RA, IL2, IL10, CXCL10, CCL2, IL1B, AGT, CXCL8, IL6, TNF, IFNG, VEGFA, IL4, ICAM1, STAT3, IL13, IL5, IL17A, CSF2, CRP, IL15, CXCL2, and MMP9 in descending order. The pathway enrichment tests were performed with selected gene sets by using an over-representation analysis approach [28]. The 10 most relevant pathways were sorted by p-value (Fig. 5b). The most significant enriched pathways were related to anti-inflammatory phenotypes and/or the human immune system.

#### Drug scoring results for prioritizing repurposable drugs

We performed a scoring algorithm to predict candidate repurposable drugs for COVID-19 with the complemented disease-gene-drug network. The initial label of graph-based SSL was set to COVID-19 only ( $\gamma_{\text{COVID}} \in \{+1\}$ ). COVID-19 was connected with 18 diseases and 17 genes directly after complementation, but none of the drugs were directly connected with COVID-19. Even though COVID-19 was not directly connected with any drugs, the candidate drugs can be predicted indirectly by considering the proximities between intra-layer relations and inter-layer relations. The scoring algorithm propagates the one positive label from COVID-19 node to remaining unlabeled node with the underlying structure of network. Since there were no approved therapeutic treatments for COVID-19 as of May 2020, there are no ground truths for the predicted results making it difficult to evaluate their accuracy in this study.

The entirety of the drugs as shown in the scoring curves could be candidates for repurposable drugs for COVID-19 (Fig. 6a). The dark to light colors in scoring curves and networks' nodes represent the normalized scores. In order to provide a list of candidate drugs with data-driven evidence, we recommended the top-30 candidate repurposable drugs from the scoring curve (normalized scores > 0.5). Figure 6b depicts the sub-network with the recommended candidates and Fig. 6c shows the detailed predicted scores. Steroids such as dexamethasone, prednisolone, and hydrocortisone were recommended as top candidates. Among them, dexamethasone, an anti-inflammatory drug, had the highest scores. Dexamethasone is a low cost steroid that reduces inflammation by mimicking anti-inflammatory hormones produced by the body. COVID-19 treatment guidelines recommend using 6 mg per day dose of dexamethasone for up to 10 days for hospitalized patients with COVID-19.

In addition, we were able to find 17 cases preparing study protocols or recruiting for clinical trials with recommended drugs as of the end of May 2020 (regardless

of whether or not they were discontinued as of September 2022, reported in <https://clinicaltrials.gov>). In addition, we searched the literature for evidence or possible relationships between the list of drugs and COVID-19 in order to investigate the potential of repositioning in the recent studies. Since the 30 candidate repurposable drugs were prioritized in the complemented network based on past time study points. Relevant studies identified for determining possible therapeutic candidate in COVID-19 can be found in the addit. (Additional file 2: Table S2).

#### Associations of prioritized repurposable drugs with COVID-19 phenotypes using electronic health records

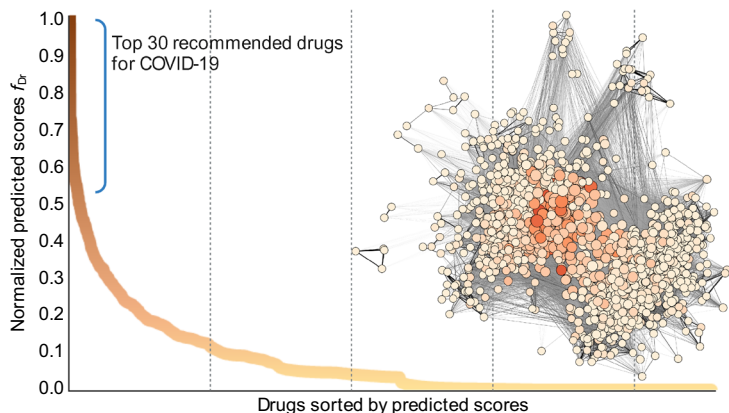
To provide additional evidence for the utility of network-based drug repurposing with complemented network, we conducted an EHR-based medication analysis using the prioritized candidate drugs and several COVID-19 phenotypes. Among ~160 K patients in the Penn Medicine COVID-19 registry, which includes all COVID-19 RT-PCR test results within the health system, as of October 2021, we extracted medication order data to investigate associations with the candidate drugs and various COVID-19 outcome phenotypes. To perform these statistical analyses, we assigned case and control status for each of three COVID-19 outcomes: (a) COVID-19 susceptibility, (b) COVID-19 positive hospital admission, and (c) COVID-19 severity.

(a) COVID-19 susceptibility was determined by a positive RT-PCR test whereas patients who had only ever had a COVID-19 negative RT-PCR test in the registry were labeled as a control. (b) A COVID-19 positive hospital admission was determined by an inpatient hospitalization with a primary diagnosis ICD-10 code U07.1 used for COVID-19 diagnosis and a positive RT-PCR test or a primary diagnosis of an ICD code indicative of a COVID-19 related symptoms, a positive RT-PCR test, and a clinician chart review for admissions dated prior to the usage of U07.1. Remaining cases from the COVID-19 susceptibility phenotype were considered as controls. (c) COVID-19 severity was determined by use of ventilator and/or an intensive care unit stay during an inpatient hospitalization as defined previously in (b) while remaining cases from the COVID-19 hospitalization phenotype were considered as controls.

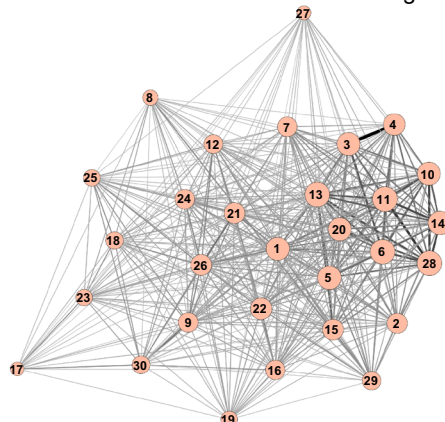
We ran logistic regressions for each outcome, using a selected candidate drug as a predictor and adjusting for age, gender, and the self-reported race as described in Methods. We only included patients with four or more encounters within the Penn Medicine hospital system prior to their COVID-19 RT-PCR test to reduce bias, i.e. patients who may have received COVID-19 related care at Penn Medicine but were not regular patients and thus would not have historical medication order data.



**a** Drug scoring curves and network for 2,713 drugs



**b** Sub-network with recommended drugs



**c** Predicted scores of top 30 candidate repurposable drugs for COVID-19

Drug (Compound)	Predicted Scores	Normalized predicted scores $f_r$					Pharmaceutical Categories	Clinical Trial (As of May 2020)
		1.0	0.9	0.8	0.7	0.6		
1 Resveratrol	1.000	○					Miscellaneous	
2 Methotrexate	0.728				○		Cytotoxic medication (chemotherapy)	✓
3 Dexamethasone	0.727				○		Steroids	✓
4 Indomethacin	0.722				○		Analgesics	✓
5 Quercetin	0.708				○		Miscellaneous	
6 Bisphenol A	0.691				○		Miscellaneous	
7 Zinc	0.678				○		Element / Vitamin / Chemical compound	✓
8 Prednisolone	0.674				○		Steroids	
9 Simvastatin	0.645				○		Lipid lowering medication	✓
10 Doxorubicin	0.629				○		Cytotoxic medication (chemotherapy)	
11 Cyclosporine	0.608				○		Immunosuppressant	✓
12 Ascorbic Acid	0.604				○		Element / Vitamin / Chemical compound	✓
13 Estradiol	0.600				○		Miscellaneous	✓
14 Acetaminophen	0.591				○		Analgesics	✓
15 Progesterone	0.585				○		Hormonal medication	✓
16 Rosiglitazone	0.576				○		Antidiabetic drug	
17 Enalapril	0.569				○		ACE inhibitor	✓
18 Dinoprostone	0.565				○		Prostaglandin	
19 Ibuprofen	0.560				○		Analgesics	✓
20 Tretinoin	0.556				○		Element / Vitamin / Chemical compound	✓
21 Curcumin	0.550				○		Miscellaneous	
22 Troglitazone	0.549				○		Antidiabetic drug	
23 Hydrocortisone	0.548				○		Steroids	✓
24 Diclofenac	0.543				○		Analgesics	
25 Nitric Oxide	0.537				○		Element / Vitamin / Chemical compound	✓
26 Acetylcysteine	0.532				○		Respiratory symptomatic treatment	✓
27 Theophylline	0.528				○		Respiratory symptomatic treatment	
28 Valproic Acid	0.517				○		Cytotoxic medication (chemotherapy)	
29 Azathioprine	0.515				○		Immunosuppressant	
30 Atorvastatin	0.514				○		Miscellaneous	✓

**Fig. 6** Candidate repurposable drugs with drug scoring curves



connectivity among heterogeneous biological entities. However, in the case of a situation where there is no information about the new disease in the previously accumulated database when a new infectious disease emerges, it is challenging to obtain inference until a new disease node like COVID-19 has a connection with other components from network-based approaches. To solve this connectivity problem, we developed the network-based complementary linkage method to overcome the deficiency of disconnections between new emerging diseases and an already-constructed network. One of the objectives of this study is to quickly provide medical or pharmaceutical experts with a curated catalog of potentially repurposable drugs, serving as a prompt for investigating their potential during a public health emergency. It is unreasonable to immediately use the results of computational drug repurposing as alternatives to existing therapeutics without any clinical verification, but the provision of prioritized candidate drugs based on evidence of knowledge can reduce the number of failures in early-stage trials.

From this point of view, it is more important to provide evidence for utilization through medication association analysis with EHR, rather than simply providing a list of candidates. However, the EHR analysis highlighted several challenges of searching for associations between medication use and clinical outcomes in observational medical records data. First, it is difficult to determine a direct cause of a particular medication towards an outcome; for example, a drug could show a strong association with a negative disease outcome simply because it is more frequently prescribed to patients with severe disease and not due to a negative effect of the drug itself. Similarly, patients can, and often are, taking multiple drugs simultaneously making it difficult to control for potential drug interaction effects. Lastly, using medication orders data relies on the assumption that patients filled the order and took the medication as prescribed. This assumption has more potential impact on inferences that require some knowledge of the patient outside of the healthcare system, such as for the COVID-19 susceptibility phenotype which looks prior to the patient's positive test, and is less of a concern for hospital admission outcomes where detailed, structured electronic health record data provide more certainty of the timing and administration of the drug. Despite these limitations, EHR remains an invaluable resource for identifying potential candidates for drug repositioning, particularly in the context of emerging new disease, and this method could further be expanded upon allowing for more sophisticated emulated clinical trials in large diverse patient populations.

Several prioritized drugs may be considered for the management of severe COVID-19 symptoms. First, dexamethasone, prednisolone, and hydrocortisone are steroid analogues, which affect immune and inflammatory functions. There have been reports suggesting that steroids may be effective in the control of systemic inflammation or 'cytokine storm' in severe COVID-19 cases [29], and there are several on-going trials on the effectiveness of steroid treatment [30]. The current study also supports the possibility of steroid therapy in patients with COVID-19. Until now, there has been controversy regarding the use of NSAIDs in COVID-19 patients. With the inclusion of acetaminophen as one of the highlighted drugs in our study, we wish to bring attention to the role NSAIDs may play in helping an individual with COVID-19. Given the body's inflammatory response to the virus, researchers have been studying the effects of some immune-modulating drugs including methotrexate and cyclosporine [31, 32], although there is a paucity of information on other immune-modulating medications or cytotoxic drugs including azathioprine, doxorubicin, valproic acid, and arsenic trioxide.

Based on the current study, further studies are needed to evaluate the possibility of immune-modulating drugs in the context of COVID-19. Several drugs also require attention, such as ACE inhibitors (enalapril), lipid lowering medications (simvastatin), hormonal medications (estradiol, progesterone) and antidiabetic drugs (rosiglitazone, troglitazone). These medications may be more effective in populations with specific comorbidities such as kidney disease, diabetes, or coronary/cardiovascular disease there is need to evaluate the efficacy of these medications in these populations. However, of course, before any of these potential treatments are given to help patients suffering from COVID-19, rigorous clinical trials are required.

## Conclusion

In this study, we developed a network-based drug repurposing method for rapid screening to respond to the situation of a new emerging disease, to present solutions and new methodologies to address the lessons learned from the COVID-19 pandemic in terms of network-based drug repurposing. We proposed the network-based complementary linkage method to overcome the deficiency of disconnections between new emerging diseases and an already-constructed network. To simulate and test our proposed method, we assumed situation as an early stage of the pandemic with insufficient information related to COVID-19. We constructed a backbone network using publicly available biomedical and pharmaceutical data and fragmented COVID-19-related information was reinforced into the backbone network by applying the

proposed complementary linkage method. To translate the complemented network for finding candidate treatments, network-based label propagations were applied and we validated the prioritized candidate drugs with EHR-based medication analysis.

There are several limitations of our studies. From the point of view of the biomedical networks, this study only investigated the heterogeneous relationships between disease, genes and drugs. However, to develop more sophisticated repurposable drugs, complex relationship information can be utilized in the network. For example, single-nucleotide polymorphisms obtained through genome-wide association studies, or green nanomaterials targeting specific cells or mitochondria can be utilized [33, 34]. Another particular limitation of our study is the relatively small sample size of the few databases we utilized; however, this concern is quickly alleviated as the robust yet flexible nature of a network-based approach allows us to very easily supplement and correct our current model. As we receive the newest information regarding the novel coronavirus, we can easily update the candidate drug/gene components of our networks, perform a set of updated calculations and generate an updated gene and drug candidate list almost instantly. With this in mind, we hope our approach may help clinicians and scientists make the difficult decisions regarding which drugs or gene targets to test first in this global race for a cure.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12967-023-04223-2>.

**Additional file 1.** Extended Methods.

**Additional file 2: Table S1.** Data for constructing backbone network.  
**Table S2.** COVID-19 relational data.

**Additional file 3.** List of gene/protein scores for COVID-19.

## Acknowledgements

This work was supported by the National Institutes of Health [R01 AG071470].

## Author contributions

YN developed the study design, developed methodology, analyzed and interpreted the data, and wrote the initial manuscript. AL analyzed electronic health record data and interpreted the results. J-S. Y, S.M. L and J.W. P, interpreted the discussion and revised the manuscript. Z. C and B. L contributed to interpreting results and revised manuscript. X. N and L. S contributed to the discussion and revised manuscript. A. V supervised the study, collected electronic health record data, implemented the validation, and revised manuscript. D. K developed the study design, supervised the study, contributed to the discussion, and revised the manuscript. All authors read and revised the manuscript. All authors read and approved the final manuscript.

## Funding

This work was supported by the National Institutes of Health [R01 AG071470].

## Availability of data and materials

Data for constructing network are publicly available at CTD (the Comparative Toxicogenomics Database, <http://ctdbase.org/>), STRING (<https://string-db.org/>), and DrugBank (<https://go.drugbank.com/>). The Penn Medicine electronic health record data cannot be shared publicly due to the violation of patient privacy and the absence of informed consent for data sharing.

## Declarations

### Ethics approval and consent to participate

The full study protocol for Penn Medicine EHR analysis was approved by the University of Pennsylvania Institutional Review Board (IRB) under the protocol for the study titled "Clinical, social, and genetic risk stratification for COVID-19 outcomes" (Protocol #844360).

### Consent for publication

Not applicable.

### Competing interests

The authors declare that there are no competing interests.

### Author details

<sup>1</sup>Department of Biostatistics, Epidemiology & Informatics, The Perelman School of Medicine, University of Pennsylvania, B304 Richards Building, 3700 Hamilton Walk, Philadelphia, PA 19104-6116, USA. <sup>2</sup>Genomics and Computational Biology Graduate Group, Perelman School of Medicine, University of Pennsylvania, Philadelphia, USA. <sup>3</sup>Division of Endocrinology and Metabolism, Department of Internal Medicine, St. Vincent's Hospital, College of Medicine, The Catholic University of Korea, Seoul, South Korea. <sup>4</sup>Department of Obstetrics and Gynecology, Seoul National University College of Medicine, Seoul, South Korea. <sup>5</sup>Department of Surgery, Seoul National University College of Medicine, Seoul, South Korea. <sup>6</sup>Computer Science and Engineering Department, College of Engineering, The Ohio State University, Columbus, USA. <sup>7</sup>Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, USA. <sup>8</sup>Biomedical Informatics Department, College of Medicine, The Ohio State University, Columbus, USA. <sup>9</sup>Translational Data Analytics Institute, The Ohio State University, Columbus, USA. <sup>10</sup>Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, USA.

Received: 20 February 2023 Accepted: 24 May 2023

Published online: 26 June 2023

## References

- Ashburn TT, Thor KB. Drug repositioning: identifying and developing new uses for existing drugs. *Nat Rev Drug Discov.* 2004;3:673–83.
- Jarada TN, Rokne JG, Alhaji R. A review of computational drug repositioning: strategies, approaches, opportunities, challenges, and directions. *J cheminform.* 2020. <https://doi.org/10.1186/s13321-020-00450-7>.
- Martinez MA. Efficacy of repurposed antiviral drugs: lessons from COVID-19. *Drug Discovery Today.* 2022. <https://doi.org/10.1016/j.drudis.2022.02.012>.
- Eastman RT, Roth JS, Brimacombe KR, Simeonov A, Shen M, Patnaik S, Hall MD. Remdesivir: a review of its discovery and development leading to emergency use authorization for treatment of COVID-19. *ACS Cent Sci.* 2020;6:672–83.
- Lammers T, Sofias AM. Dexamethasone nanomedicines for COVID-19. *Nat Nanotechnol.* 2020;15:622–4.
- Noreen S, Maqbool I, Madni A. Dexamethasone: Therapeutic potential, risks, and future projection during COVID-19 pandemic. *Eur J Pharmacol.* 2021;894: 173854.
- Singh N, Villoutreix BO. Resources and computational strategies to advance small molecule SARS-CoV-2 discovery: Lessons from the pandemic and preparing for future health crises. *Comput Struct Biotechnol J.* 2021;19:2537–48.
- Kwon OS, Kim W, Cha HJ, Lee H. In silico drug repositioning: from large-scale transcriptome data to therapeutics. *Arch Pharm Res.* 2019;42:879–89.

9. Aronsky I, Masoudi-Sobhanzadeh Y, Cappuccio A, Zaslavsky E. Advances in the computational landscape for repurposed drugs against COVID-19. *Drug Discovery Today*. 2021;26:2800–15.
10. Santos SS, Torres M, Galeano D, Sánchez MDM, Cernuzzi L, Paccanaro A. Machine learning and network medicine approaches for drug repositioning for COVID-19. *Patterns (N Y)*. 2022;3: 100396.
11. Zhou Y, Hou Y, Shen J, Huang Y, Martin W. Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell Discov*. 2020. <https://doi.org/10.1038/s41421-020-0153-3>.
12. Badkas A, De Landtsheer S, Sauter T. Topological network measures for drug repositioning. *Brief Bioinform*. 2021. <https://doi.org/10.1093/bib/bbaa357>.
13. Dotolo S, Marabotti A, Facchiano A, Tagliaferri R. A review on drug repurposing applicable to COVID-19. *Brief Bioinform*. 2021;22:726–41.
14. Nam Y, Kim M, Lee K, Shin H. CLASH: complementary linkage with anchoring and scoring for heterogeneous biomolecular and clinical data. *BMC Med Inform Decis Mak*. 2016;16(Suppl 3):72.
15. Nam Y, Kim M, Chang HS, Shin H. Drug repurposing with network reinforcement. *BMC Bioinformatics*. 2019;20:383.
16. Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson KW, Barnaby DP, Becker LB, Chelico JD, Cohen SL, et al. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City Area. *JAMA*. 2020;323:2052–9.
17. Nam Y, Jhee JH, Cho J, Lee JH, Shin H. Disease gene identification based on generic and disease-specific genome networks. *Bioinformatics*. 2019;35:1923–30.
18. Subramanya A, Talukdar PP. Graph-based semi-supervised learning. *Synth Lect Artif Intell Mach Learn*. 2014;8:1–125.
19. Chong Y, Ding Y, Yan Q, Pan S. Graph-based semi-supervised learning: a review. *Neurocomputing*. 2020;408:216–30.
20. Kim M, Nam Y, Shin H. An inference method from multi-layered structure of biomedical data. *BMC Med Inform Decis Mak*. 2017;17:52.
21. Org C-HGlab. The COVID-19 host genetics initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. *European J Hum Genet*. 2020;28:715–8.
22. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, Chang Z, Woolsey J. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res*. 2006;34:D668–672.
23. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47:D607–d613.
24. Davis AP, Grondin CJ, Johnson RJ, Sciaky D, McMoran R, Wieggers J, Wieggers TC, Mattingly CJ. The comparative toxicogenomics database: update 2019. *Nucleic Acids Res*. 2019;47:D948–d954.
25. Li B, Yang J, Zhao F, Zhi L, Wang X, Liu L, Bi Z, Zhao Y. Prevalence and impact of cardiovascular metabolic diseases on COVID-19 in China. *Clin Res Cardiol*. 2020;109:531–8.
26. Wang B, Li R, Lu Z, Huang Y. Does comorbidity increase the risk of patients with COVID-19: evidence from meta-analysis. *Aging (Albany NY)*. 2020;12:6049–57.
27. De Meo P, Ferrara E, Fiumara G, Provetti A. Generalized louvain method for community detection in large networks. In 2011 11th international conference on intelligent systems design and applications. IEEE; 2011: 88–93.
28. Gillespie M, Jassal B, Stephan R, Milacic M, Rothfels K, Senff-Ribeiro A, Griss J, Sevilla C, Matthews L, Gong C, et al. The reactome pathway knowledge-base 2022. *Nucleic Acids Res*. 2022;2022(50):D687–d692. <https://doi.org/10.1093/nar/gkab1028>.
29. Ledford H. Coronavirus breakthrough: dexamethasone is first drug shown to save lives. *Nature*. 2020;582:469.
30. Lythgoe MP, Middleton P. Ongoing clinical trials for the management of the COVID-19 pandemic. *Trends Pharmacol Sci*. 2020;41:363–82.
31. Frohman EM, Villemarette-Pittman NR, Cruz RA, Longmuir R, Rowe V, Rowe ES, Varkey TC, Steinman L, Zamvil SS, Frohman TC. Part II. high-dose methotrexate with leucovorin rescue for severe COVID-19: an immune stabilization strategy for SARS-CoV-2 induced “PANIC” attack. *J Neurol Sci*. 2020;415:116935.
32. Glowacka P, Rudnicka L. The antiviral properties of cyclosporine. Focus on coronavirus, hepatitis C virus, influenza virus, and human immunodeficiency virus infections. *Biology*. 2020. <https://doi.org/10.3390/biology9080192>.
33. Ramazanli VN, Ahmadov IS. Synthesis of silver nanoparticles by using extract of olive leaves. *Adv Biol Earth Sci* 2022;7(3):238–244. <https://doi.org/10.1016/j.ijbiomac.2023.124924>.
34. Baran A, Firat Baran M, Keskin C, Hatipoğlu A, Yavuz Ö, İrtegin Kandemir S, Adican MT, Khalilov R, Mammadova A, Ahmadian E. Investigation of antimicrobial and cytotoxic properties and specification of silver nanoparticles (AgNPs) derived from *Cicer arietinum* L. green leaf extract. *Front Bioeng Biotechnol*. 2022;10:263.

## Publisher's Note

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

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

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

At BMC, research is always in progress.

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

